

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:01:28 ; Search time 5.29303 Seconds
(without alignments)
817.488 Million cell updates/sec

Title: US-09-986-682B-1
Perfect score: 114
Sequence: 1 MNSGDYKEDYGFATHTRADML 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_protist:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	100.0	487	2	082854
2	59	51.8	489	16	097181
3	55	48.2	487	10	093290
4	55	48.2	646	10	093290
5	52	45.6	524	3	09P330
6	50	43.9	220	5	09Y0A8
7	48	42.1	478	16	09KRS8
8	48	42.1	499	2	09Z5F5
9	48	42.1	582	4	08W1W3
10	48	42.1	670	4	09BRR2
11	48	42.1	673	4	09BY02
12	48	42.1	832	11	09CX82
13	47	41.2	289	10	09SZJ8
14	47	41.2	435	10	065620
15	47	41.2	567	16	098R03
16	46	40.4	365	17	059328

17	46	40.4	677	10	09Z051	09Z051 arabidopsis
18	45	39.5	201	5	09VTV3	09VTV3 drosophila
19	45	39.5	526	3	096TV7	096TV7 venturia na
20	45	39.5	608	16	09HZV4	09HZV4 pseudomonas
21	45	39.5	649	13	09DDC0	09DDC0 pleurodeles
22	45	39.5	728	10	004512	004512 arabidopsis
23	45	39.5	729	10	09ZU22	09ZU22 arabidopsis
24	45	39.5	749	10	09C6M5	09C6M5 arabidopsis
25	45	39.5	882	10	065231	065231 arabidopsis
26	45	39.5	893	10	09LHT1	09LHT1 arabidopsis
27	45	39.5	895	10	080973	080973 arabidopsis
28	45	39.5	941	10	09LTH0	09LTH0 arabidopsis
29	45	39.5	942	10	09SLI8	09SLI8 arabidopsis
30	45	39.5	942	10	09ZU08	09ZU08 arabidopsis
31	45	39.5	943	10	09SPY7	09SPY7 arabidopsis
32	45	39.5	1071	10	09MAK2	09MAK2 arabidopsis
33	45	39.5	1127	10	09C608	09C608 arabidopsis
34	45	39.5	6781	12	091AV2	091AV2 porcine epi
35	44	38.6	210	2	08RIW9	08RIW9 campylobact
36	44	38.6	250	16	09A735	09A735 caulobacter
37	44	38.6	253	12	09PXJ7	09PXJ7 transmissib
38	44	38.6	342	4	09HCA5	09HCA5 homo sapien
39	44	38.6	364	5	021916	021916 caenorhabdi
40	44	38.6	365	3	09HDU4	09HDU4 schizosacch
41	44	38.6	392	3	096WU4	096WU4 zygosacchar
42	44	38.6	459	17	08U3P5	08U3P5 pyrococcus
43	44	38.6	484	16	09PGI8	09PGI8 xylella fas
44	44	38.6	600	3	09HG12	09HG12 neurospora
45	44	38.6	677	13	090XA5	090XA5 cyprinus ca

ALIGNMENTS

RESULT 1

ID	082854	PRELIMINARY:	PRT:	487 AA.
AC	082854;			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DR	01-NOV-1998 (TREMBLrel. 08, last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)			
DE	Beta-fructofuranosidase.			
OS	Bacillus sp.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;			
OC	Bacillaceae; Bacillus.			
OX	NCBI_TaxID-1409;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-V230;			
RA	Tsutsaki K.;			
RT	Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.			
RT	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-V230;			
RA	Kurimoto M.;			
RT	"Cloning and sequencing of beta-fructofuranosidase gene from Bacillus			
RT	sp. V230";			
RL	Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AB010272; BAA32083.1; -			
DR	InterPro; IPR003469; Glyco_hydro_68.			
DR	Pfam; PF02435; Glyco_hydro_68; 1.			
SQ	SEQUENCE 487 AA; 53412 MW; 7F52A47921824AE3 CRC64;			

Query Match 100.0%; Score 114; DB 2; Length 487;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFATHTRADML 21
|||||
DB 33 MNSGDYKEDYGFATHTRADML 53

RESULT 2

```

097181
ID 097181 PRELIMINARY: PRT: 489 AA.
AC 097181:
DT 01-OCT-2001 (TREMblrel. 18, Created)
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Levanusucrase.
GN CAC1772.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hiltl J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007686; AAK79737.1; -
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
KW Complete proteome.
SQ SEQUENCE 489 AA: 5453 MW: 9002CB36AFD3CD8 CRC64:

Query Match 51.8%; Score 59; DB 16; Length 489;
Best Local Similarity 57.9%; Pred. No. 0.26;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 NSGDYKEDYGFHATIRAD 20
DB 30 NDMNKETGYSHIRIM 48

RESULT 3
093290
ID 093290 PRELIMINARY: PRT: 487 AA.
AC 093290:
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE AT4924430/T22A6.260.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Caninici P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY057715; AAL5345.1; -
SQ SEQUENCE 487 AA: 5538 MW: 1884FED01571F262 CRC64:

Query Match 48.2%; Score 55; DB 10; Length 487;
Best Local Similarity 69.2%; Pred. No. 1.2;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 5 DYKEDYGFHATIR 17
DB 347 DYKDDWFFAHVTR 359

```

```

RESULT 4
09STV1
ID 09STV1 PRELIMINARY: PRT: 646 AA.
AC 09STV1:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE LG27/30-like gene.
GN T22A6.260 OR AT4624430.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Zimmermann W., Gruenisen A., Wambutt R., Bancroft I.,
RA Meves H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Zimmermann W., Gruenisen A., Wambutt R., Kalicki J., Wohlmann P.,
RA Smith A., Meves H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL078637; CAB45078.1; -
DR EMBL: AL161561; CAB79353.1; -
SQ SEQUENCE 646 AA: 73608 MW: B5ACAD2021EAE0C3 CRC64:

Query Match 48.2%; Score 55; DB 10; Length 646;
Best Local Similarity 69.2%; Pred. No. 1.6;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 5 DYKEDYGFHATIR 17
DB 506 DYKDDWFFAHVTR 518

RESULT 5
09P330
ID 09P330 PRELIMINARY: PRT: 524 AA.
AC 09P330:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE 14 alpha-demethylase (14-a demethylase).
GN CYP51A1.
OS Venturia inaequalis.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
OC Dothideomycetes et Chaetothyrionmycetes incertae sedis; Venturiaceae;
OC Venturia.
OX NCBI_TaxID=5025;
RN [1]
RP SEQUENCE FROM N.A.
RA Schinabel G., Jones A.L.;
RC STRAIN-ENT27;
RT "Isolation and molecular characterization of the 14-a demethylase gene
RT from Venturia inaequalis and its polymorphic 5' flanking region.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ENT54;
RA Schinabel G., Jones A.L.;
RT "Isolation and molecular characterization of the 14-a demethylase gene

```

RT of *Venturia inaequalis* and its polymorphic 5' flanking region."
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-F445;
RA Schnabel G., Jones A.L.;
RT "Isolation and molecular characterization of the 14-a demethylase gene
from *Venturia inaequalis* and its polymorphic 5' flanking region."
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AF227920: AAF76464.1: -;
DR EMBL: AF262756: AAF71293.1: -;
DR EMBL: AF262757: AAF71294.1: -;
DR HSSP: P77901: 1E9X.
DR InterPro: IPR001128: Cytochrome_P450.
DR Pfam: PF00067: P450.1.
DR PRINTS: PR00385: P450.1.
DR PROSITE: PS00086: CYTOCHROME_P450; UNKNOWN_1.
DR Heme: Methyltransferase; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 524 AA: 59075 MW: AACFE2A1CD7A84B3 CRC64;

Query Match 45.6%; Score 52; DB 3; Length 524;
Best Local Similarity 56.2%; Pred. No. 4;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 NSGDYKEDYGFAMITR 17
DB 435 NGGEKEKDYGYGLIRK 450

RESULT 6
O9Y0A8 PRELIMINARY; PRT: 220 AA.
ID O9Y0A8:
AC O9Y0A8:
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Type 2 adenylate kinase.
OS *Cryptosporidium parvum*.
OC *Cryptosporidium parvum*.
OC Eukaryota: Alveolata: Apicomplexa: Coccidia: Elmeriida:
OC Cryptosporididae: Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KSU-1;
RA Riordan C.E., Langreth S.G., Sanchez L.B., Upton S.J., Kayser O.,
RA Zhu G., Keithly J.S.;
RT "Preliminary Evidence for a Mitochondrion in *Cryptosporidium parvum*:
Phylogenetic and Therapeutic Implications."
RL J. Eukaryot. Microbiol. 0:0-0(1999).
CC -1- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
DR EMBL: AF163320: AAD45371.1: -;
DR HSSP: P07170: 1AKY.
DR InterPro: IPR000850: Adenylate_kin.
DR Pfam: PF00406: adenylatekinase.1.
DR PRINTS: PR00094: ADENYLATE_KINASE.
DR Prodom: PD000657: Adenylate_kin.1.
DR PROSITE: PS00113: ADENYLATE_KINASE.1.
KW kinase; Transferase
SQ SEQUENCE 220 AA: 24228 MW: CAID6999A68A6DF CRC64;

Query Match 43.9%; Score 50; DB 5; Length 220;
Best Local Similarity 53.3%; Pred. No. 3.2;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 7 KEDIGFAHITRADML 21
DB 27 KKEYGLAHSTGDM 41

RESULT 7
O9KRS8 PRELIMINARY; PRT: 478 AA.
ID O9KRS8

AC O9KRS8:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE 6-phospho-beta-glucosidase.
GN VC1558.
OS *Vibrio cholerae*.
OC Bacteria: Proteobacteria: gamma subdivision: Vibrionaceae: *Vibrio*.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Yamathayan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ulteback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
cholerae".
RT Nature 406:477-483(2000).
RL Nature 406:477-483(2000).
DR EMBL: AE004233: AAF94712.1: -;
DR HSSP: P1546: 1PBG.
DR TIGR: VC1558;
DR InterPro: IPR001360: GH_1.
DR Pfam: PF00232: Glyco_hydro_1.1.
DR PRINTS: PR00131: GLYHYDRLASE1.
DR Prodom: PD000650: GH_1.1.
DR PROSITE: PS00572: GLYCOSYL_HYDROL_FL_1; UNKNOWN_1.
DR PROSITE: PS00653: GLYCOSYL_HYDROL_FL_2; 1.
KW Complete proteome.
SQ SEQUENCE 478 AA: 55214 MW: DC7642B983300F87 CRC64;

Query Match 42.1%; Score 48; DB 16; Length 478;
Best Local Similarity 41.2%; Pred. No. 17;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 3 SGDKYEDYGFAMITRAD 19
DB 434 TGEYKRRYGFIVYDKD 450

RESULT 8
O9Z5E5 PRELIMINARY; PRT: 499 AA.
ID O9Z5E5:
AC O9Z5E5:
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE *Levansucrase*.
GN SABC.
OS *Paenibacillus polymyxa* (*Bacillus polymyxa*).
OC Bacteria: Firmicutes: *Bacillus*/Clostridium group: Bacillales;
OC Paenibacillaceae: *Paenibacillus*.
OX NCBI_TaxID=1406;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CF43;
RX MEDLINE=21042000; PubMed=11200435;
RA Bezzate S., Aymerich S., Chamber R., Czarne S., Berge O., Heulin T.;
RT "Disruption of the *Paenibacillus polymyxa* *levansucrase* gene impairs
its ability to aggregate soil in the wheat rhizosphere."
RL Environ. Microbiol. 2:333-342(2000).
DR EMBL: AJ133737: CAB93327.1: -;
DR InterPro: IPR003469: Glyco_hydro_68.
DR Pfam: PF02435: Glyco_hydro_68.1.
SQ SEQUENCE 499 AA: 55317 MW: E646CD986292336C CRC64;

Query Match 42.1%; Score 48; DB 2; Length 499;
Best Local Similarity 50.0%; Pred. No. 17;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY 2 NSGPKEDYGFPHITRADML 21
 :: | | | | | | | | | |
 DB 33 DASDKTDNAFTQITRNDML 52

RESULT 9

08MYW3 PRELIMINARY: PRT: 582 AA.
 AC 08MYW3.
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE Hypothetical 66.3 kDa protein.
 GN p13624.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhang P.P., Zhou X.M., Jiang H.Q., Huang Y., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF318335; AAL55842.1;
 KW Hypothetical protein.
 SQ SEQUENCE 582 AA; 66322 MW; B847F093D01B1982 CRC64;

Query Match 42.1%; Score 48; DB 4; Length 582;
 Best Local Similarity 31.2%; Pred. No. 21;
 Matches 10; Conservative 4; Mismatches 2; Indels 16; Gaps 1;

OY 5 DYKEDYG-----FAHITRADM 20
 ||:| | |
 DB 297 DYEEKYGRMPREWCMAERIAVEFCHVTRAEI 328

RESULT 10

09BRR2 PRELIMINARY: PRT: 670 AA.
 AC 09BRR2.
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Similar to RIKEN cDNA 3100002B05 gene.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE-PLACENTA;
 RC Strausberg R.;
 RA Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC006116; AAH06116.1;
 SQ SEQUENCE 670 AA; 76351 MW; 8057AAC082136EC7 CRC64;

Query Match 42.1%; Score 48; DB 4; Length 670;
 Best Local Similarity 31.2%; Pred. No. 24;
 Matches 10; Conservative 4; Mismatches 2; Indels 16; Gaps 1;

OY 5 DYKEDYG-----FAHITRADM 20
 ||:| | |
 DB 268 DYEEKYGRMPREWCMAERIAVEFCHVTRAEI 299

RESULT 11

09BYO2 PRELIMINARY: PRT: 673 AA.
 AC 09BYO2.
 DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 76.7 kDa protein.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhao X., Wan D., Han L., Gu J.;
 RT "cDNA clone C63R on chromosome 17p13.3."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF246287; AAK27973.1;
 KW Hypothetical protein.
 SQ SEQUENCE 673 AA; 76667 MW; 8257C0D59E9385C1 CRC64;

Query Match 42.1%; Score 48; DB 4; Length 673;
 Best Local Similarity 31.2%; Pred. No. 24;
 Matches 10; Conservative 4; Mismatches 2; Indels 16; Gaps 1;

OY 5 DYKEDYG-----FAHITRADM 20
 ||:| | |
 DB 297 DYEEKYGRMPREWCMAERIAVEFCHVTRAEI 328

RESULT 12

09CX82 PRELIMINARY: PRT: 832 AA.
 AC 09CX82.
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 3100002B05RIK protein.
 GN 3100002B05RIK
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arawaka T., Hara A., Fukunishi Y., Konno H., Aachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
 RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann W., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
 RA Suzuki H., Toyokawa K., Wang K.H., Welter C., Whitaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK019408; BAB31707.1;
 DR MGD: MGI:1915549; 3100002B05RIK.
 SQ SEQUENCE 832 AA; 94464 MW; C22B4496491ARCDC CRC64;

Query Match 42.1%; Score 48; DB 11; Length 832;
 Best Local Similarity 31.2%; Pred. No. 31;
 Matches 10; Conservative 4; Mismatches 2; Indels 16; Gaps 1;

OY 5 DYKEDYG-----FAHITRADM 20
 ||:| | |
 DB 297 DYEEKYGRMPREWCMAERIAVEFCHVTRAEI 328

RESULT 13
ID 095ZJ8 PRELIMINARY; PRT: 289 AA.
AC 095ZJ8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Hypothetical 33.4 kDa protein.
GN F20D10.80 OR ATAG37960.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Beyer M., Wedler H., Kutzner M., Wambutt R., Bancroft I., Newes H.W.,
Mayer K.F.X., Schueller C.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Wedler H., Kutzner M., Wambutt R., Mewes H.W., Lemcke K.,
Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL035538; CAB37536.1; -
DR EMBL: AL161592; CAB80461.1; -
KW Hypothetical protein.
SQ SEQUENCE 289 AA; 3385 MW; C0BFEE6AF7EEF554 CRC64;
SO
Query Match 41.2%; Score 47; DB 10; Length 289;
Best Local Similarity 53.8%; Pred. No. 14;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
OY 5 DYKEDYGFANITR 17
Db 157 DYKDMFYAHVNR 169
RESULT 14
ID 065620 PRELIMINARY; PRT: 435 AA.
AC 065620;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MYST1 protein (Fragment).
GN MYST1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=COLUMBIA, AND CV. CO;
RX MEDLINE=20267440; PubMed=10809443;
RA Tavares R., Aubourg S., Decharny A., Kreis M.;
RT "Organization and structural evolution of four multigene families in
Arabidopsis thaliana: ATLCA, ATLGT, ATWST and ATHD-GL2.";
RL Plant Mol. Biol. 42:703-717(2000).
DR EMBL: Y16848; CAA76417.1; -
FT NON_TER 1
SQ SEQUENCE 435 AA; 49882 MW; 9591F93B67214562 CRC64;

Query Match 41.2%; Score 47; DB 10; Length 435;
Best Local Similarity 53.8%; Pred. No. 22;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
OY 5 DYKEDYGFANITR 17
Db 303 DYKDMFYAHVNR 315
RESULT 15
ID 098R03 PRELIMINARY; PRT: 567 AA.
AC 098R03;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE LIPOPROTEIN.
GN MYP2_2070.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Hellig R., Ferris S., Barbe V., Samson D., Galisson F.,
Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
Mycoplasma pulmonis.";
RL Nucleic Acids Res. 29:2145-2153(2001).
DR EMBL: AL445563; CAC13380.1; -
DR Mypulast; MYP2_2070; -
DR InterPro: IPR002057; Isopen_N_synth.
DR PROSITE: PS00186; IPNS_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 567 AA; 64792 MW; C6D953B917B76574 CRC64;
SO
Query Match 41.2%; Score 47; DB 16; Length 567;
Best Local Similarity 44.4%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
OY 2 NSGDYKEDYGFANITRAD 19
Db 435 HGNKWDYVFTHTNRTN 452

Search completed: June 16, 2003, 12:09:34
Job time : 7.29303 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:56:18 ; Search time 1.33402 Seconds
(without alignments)
652.918 Million cell updates/sec

Title: US-09-986-682B-1
Perfect score: 114
Sequence: 1 MNSGDYKEDYGFHITRADML 21

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries.

Database : Swissprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	59.6	473	1 SABC_BACST	P94468 bacillus st
2	68	59.6	473	1 SABC_BACST	P05655 bacillus su
3	66	57.9	472	1 SABC_BACAM	P21130 bacillus am
4	48	42.1	295	1 MURB_RICCN	O9218 rickettsia
5	48	42.1	295	1 MURB_RICPR	O9257 rickettsia
6	46	40.4	217	1 KAD_BACSU	P16304 bacillus su
7	44	38.6	342	1 CCR6_HUMAN	O00574 homo sapien
8	44	38.6	342	1 CCR6_HUMAN	O00574 homo sapien
9	44	38.6	342	1 CCR6_HUMAN	O00574 homo sapien
10	44	38.6	342	1 CCR6_HUMAN	O00574 homo sapien
11	43	37.7	212	1 KAD_STRPY	P18457 porcine tra
12	43	37.7	212	1 KAD_STRPY	P18457 porcine tra
13	43	37.7	212	1 KAD_STRPY	P18457 porcine tra
14	43	37.7	212	1 KAD_STRPY	P18457 porcine tra
15	42.5	37.3	587	1 COB8_ONCMY	O93847 haemophilus
16	42	36.8	125	1 RNP4_CLOPE	O8xh26 clostridium
17	42	36.8	125	1 RNP4_CLOPE	O8xh26 clostridium
18	42	36.8	125	1 RNP4_CLOPE	O8xh26 clostridium
19	42	36.8	125	1 RNP4_CLOPE	O8xh26 clostridium
20	42	36.8	125	1 RNP4_CLOPE	O8xh26 clostridium
21	42	36.8	125	1 RNP4_CLOPE	O8xh26 clostridium
22	42	36.8	125	1 RNP4_CLOPE	O8xh26 clostridium
23	42	36.8	125	1 RNP4_CLOPE	O8xh26 clostridium
24	42	36.8	125	1 RNP4_CLOPE	O8xh26 clostridium
25	41.5	36.4	343	1 GUNC_CLOSF	P23340 clostridium
26	41.5	36.4	343	1 GUNC_CLOSF	P23340 clostridium
27	41	36.0	231	1 KAD5_HUMAN	O9464 homo sapien
28	41	36.0	231	1 KAD5_HUMAN	O9464 homo sapien
29	41	36.0	231	1 KAD5_HUMAN	O9464 homo sapien
30	41	36.0	231	1 KAD5_HUMAN	O9464 homo sapien
31	41	36.0	231	1 KAD5_HUMAN	O9464 homo sapien
32	40.5	35.5	504	1 GUNZ_CLOS	P23659 clostridium
33	40	35.1	130	1 LYCK_SHEEP	P80190 ovine aries

34	40	35.1	279	1 HEM6_RICPR	O9286 rickettsia
35	40	35.1	295	1 YVCI_BACSU	O06973 bacillus su
36	40	35.1	358	1 A2HS_CAVPO	O70159 cavia porce
37	40	35.1	374	1 G3PA_CHURE	P50362 chlamydomon
38	40	35.1	465	1 YHUA_ECOLI	P37197 escherichia
39	40	35.1	504	1 RA52_YEAST	P28249 saccharomyc
40	40	35.1	617	1 ASMA_ECOLI	P28249 escherichia
41	40	35.1	920	1 VGLB_HSVSM	O04464 herpesvirus
42	40	35.1	1297	1 PUR4_PASMU	O9c1w4 pasteurilla
43	40	35.1	3415	1 POLG_POWYL	O04538 t genome po
44	39.5	34.6	340	1 ARCC_AOUAE	O67724 aquilex aeo
45	39.5	34.6	933	1 VGLB_HSV1	O04463 herpesvirus

ALIGNMENTS

RESULT 1					
ID	SABC_BACST	STANDARD:	PRT:	473 AA.	
AC	P94468:				
DT	15-DEC-1998 (Rel. 37, Last sequence update)				
DT	15-DEC-1998 (Rel. 37, Last annotation update)				
DE	Levansucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl				
DE	transferase) (Sucrose 6-fructosyl transferase).				
GN	SABC OR SUBR.				
OS	Bacillus steaerothermophilus.				
OC	Bacteria; Firmicutes; Bacillales; Geobacillus.				
OX	NCBI_TaxID=1422;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 12980;				
RX	MEDLINE=98007870; PubMed=9349714;				
RA	Li Y., Triccas J.A., Ferenci T.;				
RT	"A novel levansucrase-levanase gene cluster in Bacillus				
RT	steaerothermophilus ATCC12980.";				
RL	Biochim. Biophys. Acta 1353:203-208(1997).				
CC	- i- CATALYTIC ACTIVITY: Sucrose + ((2,6)-beta-D-fructosyl)(N) =				
CC	glucose + ((2,6)-beta-D-fructosyl)(N+1).				
CC	- i- SUBCELLULAR LOCATION: Secreted.				
CC	- i- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.				
CC	- i- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
DR	EMBL: U34874; AAB97111.1;				
DR	InterPro: IPR003469; Glyco_hydro.68.				
DR	Pfam: PF02435; Glyco_hydro.68; 1.				
KW	Transferase; Glycosyltransferase; Signal.				
FT	SIGNAL 1 29				
FT	CHAIN 30 473				
FT	SEQUENCE 473 AA: 52854 MW: 12878272EC2B771E CRC64:				
QY	Query Match	59.6%; Score 68; DB 1; Length 473;			
QY	Best Local Similarity	65.0%; Pred. No. 0.0023;			
QY	Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;				
DB	2 NSGDYKEDYGFHITRADML 21				
DB	33 NOKPYKTYGISHTRHDML 52				
RESULT 2					
ID	SABC_BACSU	STANDARD:	PRT:	473 AA.	
AC	P05655; P70984;				
DT	01-NOV-1988 (Rel. 09, Created)				

DT 01-NOV-1988 (rel. 09, last sequence update)
 DT 15-JUN-2002 (rel. 41, last annotation update)
 DE Levanisucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
 transferase) (Sucrose 6-fructosyl transferase).
 GN SABC.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168.
 RA MEDLINE=85295507; PubMed=2993818;
 RA Steinmetz M., Le Coq D., Aymerich S., Gonzy-Treboul G., Gay P.;
 RT "The DNA sequence of the gene for the secreted Bacillus subtilis
 enzyme levanisucrase and its genetic control sites."
 RT Mol. Gen. Genet. 200;220-228(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Denizot F.;
 RN Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcherdt S.,
 RA Bortiss R., Bourcier L., Brans A., Braun M., Briganti S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Comerion I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Erington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel K., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Mestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,
 RA Prescan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni K.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis."
 RT Nature 390;249-256(1997).
 RN [4]
 RP SEQUENCE OF 1-62 FROM N.A.
 RC STRAIN=168 / PY79;
 RA MEDLINE=84178454; PubMed=6424671;
 RA Fouet A., Arnaud M., Klier A., Rapoport G.;
 RT "Characterization of the precursor form of the exocellular
 levanisucrase from Bacillus subtilis."
 RT Biochem. Biophys. Res. Commun. 119;795-800(1984).
 RN [5]
 RP SEQUENCE OF 1-68 FROM N.A.
 RC MEDLINE=87008406; PubMed=2428811;
 RA Shimotsu H., Henner D.J.;
 RT "Modulation of Bacillus subtilis levanisucrase gene expression by
 sucrose and regulation of the steady-state mRNA level by sacY and
 RT sacZ genes."
 RT J. Bacteriol. 168;380-388(1986).
 CC -i- CATALYTIC ACTIVITY: Sucrose + [(2,6)-beta-D-fructosyl](N) =

CC glucose + [(2,6)-beta-D-fructosyl](N+1).
 CC -i- SUBCELLULAR LOCATION: Secreted.
 CC -i- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M14202; AAA22725.1; -
 CC EMBL: 294043; CAB08015.1; -
 CC EMBL: 299121; CAB15450.1; -
 CC EMBL: K01987; AAA22724.1; -
 CC EMBL: X02730; CAA26513.1; -
 CC PIR: A25040; A25040.
 CC PIR: S07309; S07309.
 CC Subtilisin; BG10388; sabc.
 CC InterPro: IPR003469; Glyco_hydro_68.
 CC Pfam: PF02435; Glyco_hydro_68; 1.
 CC Transferrase; Glycosyltransferase; Signal; Complete proteome.
 CC FT SIGNAL 1 29
 CC CHAIN 30 473 LEVANSUCRASE.
 CC CONFLICT 12 12 V -> I (IN REF. 4).
 CC SEQUENCE 473 AA: 52971 MM: 38FP2F571B14D5B0 CRC64:
 CC
 CC Query Match 59.6%; Score 68; DB 1; Length 473;
 CC Best Local Similarity 65.0%; Pred. No. 0.0023;
 CC Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 CC
 CC OY 2 NSGDVKEEYGFNHTRADML 21
 CC DB 33 NQPKYKRYGSHITRDMML 52
 CC
 CC RESULT 3
 CC SABC_BACAM
 CC ID SABC_BACAM STANDARD; PRT; 472 AA.
 CC AC P21130.
 CC DT 01-FEB-1991 (rel. 17, Created)
 CC DT 01-FEB-1991 (rel. 17, Last sequence update)
 CC DT 15-DEC-1998 (rel. 37, Last annotation update)
 CC DE Levanisucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
 DE transferase) (Sucrose 6-fructosyl transferase).
 GN SABC.
 OS Bacillus amyloliquefaciens.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1390;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 23844;
 RA MEDLINE=91092506; PubMed=2265762;
 RA Tang L.B., Lenstra R., Borchert T.V., Vasantha N.;
 RT "Isolation and characterization of levanisucrase-encoding gene from
 RT Bacillus amyloliquefaciens."
 RT Gene 96;89-93(1990).
 RN [2]
 RP -i- CATALYTIC ACTIVITY: Sucrose + [(2,6)-beta-D-fructosyl](N) =
 CC glucose + [(2,6)-beta-D-fructosyl](N+1).
 CC -i- SUBCELLULAR LOCATION: Secreted.
 CC -i- INDUCTION: BY SUCROSE.
 CC -i- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X52988; CAA37179.1; -

[1] SEQUENCE FROM N.A.
 RP MEDLINE=90292990; PubMed=2113521;
 RA Nakamura K., Nakamura A., Takamatsu H., Yoshikawa H., Yamane K.;
 RT "Cloning and characterization of a Bacillus subtilis gene homologous
 to E. coli secY.";
 RL J. Biochem. 107:603-607(1990).
 (2)
 RN SEQUENCE FROM N.A.
 RP STRAIN=168 / Marburg;
 RC MEDLINE=96186897; PubMed=8635744;
 RA Sun J.W., Boylan S.A., Oh S.H., Price C.W.;
 RT "Genetic and transcriptional organization of the Bacillus subtilis
 spc-alpha-region.";
 RL Gene 169:17-23(1996).
 (3)
 RN SEQUENCE OF 1-116 FROM N.A.
 RX MEDLINE=90221911; PubMed=2139212;
 RA Yoshikawa H., Doi R.H.;
 RT "Sequence of the Bacillus subtilis spectinomycin resistance gene
 region.";
 RL Nucleic Acids Res. 18:1647-1647(1990).
 (4)
 RN SEQUENCE OF 1-99 FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=90251170; PubMed=2110998;
 RA Sun J.W., Boylan S.A., Thomas S.M., Dolan K.M., Oliver D.B.,
 RA Price C.W.;
 RT "Isolation of a secY homologue from Bacillus subtilis: evidence for a
 common protein export pathway in eubacteria.";
 RL Mol. Microbiol. 4:305-314(1990).
 (5)
 RN SEQUENCE OF 1-20.
 RP STRAIN=168 / ISS8;
 RC MEDLINE=97443988; PubMed=9298659;
 RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
 RA Hecker M.;
 RT "First steps from a two-dimensional protein index towards a response-
 RT regulation map for Bacillus subtilis.";
 RL Electrophoresis 18:1451-1463(1997).
 CC -1- FUNCTION: THIS SMALL UBIQUITOUS ENZYME IS ESSENTIAL FOR
 MAINTENANCE AND CELL GROWTH.
 CC -1- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- INDUCTION: BY SUPEROXIDE.
 CC -1- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 CC EMBL: D00619; BAA00496.1; -
 DR EMBL: L47971; AAB06820.1; -
 DR EMBL: M31102; AAB59119.1; -
 DR EMBL: X51329; CA35713.1; -
 DR EMBL: Z99104; CAB1913.1; -
 DR PIR: J50492; J50492.
 DR PIR: S08630; S08630.
 DR PIR: S12684; S12684.
 DR HSSP: P27142; 12IN.
 DR Subtilist; BG10446; adk.
 DR InterPro: IPR000850; Adenylate_kin.
 DR Pfam: PF00406; adenylatekinase; 1.
 DR PRINTS: PR00094; ADENYLTKINASE.
 DR ProDom: PD000657; Adenylate_kin; 1.
 DR PROSITE: PS00113; ADENYLATE_KINASE; 1.
 KW Transferase; kinase; ATP-binding; Complete proteome.
 FT NP_BIND 7 15 ATP (BY SIMILARITY).

SO SEQUENCE 217 AA; 24119 MW; ECB9ECFA4F26A1E90 CRC64;
 Query Match 40.4%; Score 46; DB 1; Length 217;
 Best Local Similarity 61.5%; Pred. No. 3.3;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 8 EDYGFHITTRADM 20
 1111 11: 11
 DB 22 EDYGFHITSTGDM 34
 RESULT 7
 CCR6_HUMAN
 ID CCR6_HUMAN STANDARD; PRT; 342 AA.
 AC 000574; 000575;
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE C-X-C chemokine receptor type 6 (CCR6) (CCR6-6) (G protein-coupled
 DE receptor bonzo) (G protein-coupled receptor STRL33).
 GN CCR6 OR BONZO OR STRL33 OR TYMSR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97373958; PubMed=9230441;
 RA Deng H.K., Unutmaz D., Kewalramani V.N., Littman D.R.;
 RT "Expression cloning of new receptors used by simian and human
 RT immunodeficiency viruses.";
 RL Nature 388:296-300(1997).
 (2)
 RN SEQUENCE FROM N.A.
 RP MEDLINE=97311099; PubMed=9166430;
 RA Liao F., Alkhalib G., Peden K.W.C., Sharma G., Berger E.A.,
 RA Farber J.M.;
 RT "STRL33, A novel chemokine receptor-like protein, functions as a
 RT fusion cofactor for both macrophage-tropic and T cell line-tropic
 RT HIV-1.";
 RL J. Exp. Med. 185:2015-2023(1997).
 (3)
 RN SEQUENCE FROM N.A.
 RP TISSUE=Blood;
 RC MEDLINE=97431687; PubMed=9285716;
 RA Loeschner M., Amara A., Oberlin E., Brass N., Legler D.F.,
 RA Loeschner P., D'Apuzzo M., Meese E.U., Rousset D., Virelizier J.L.,
 RA Baggiolini M., Arenzana-Seisdedos F., Moser B.;
 RT "TYMSR, a putative chemokine receptor selectively expressed in
 RT activated T cells, exhibits HIV-1 coreceptor function.";
 RL Curr. Biol. 7:652-660(1997).
 CC -1- FUNCTION: RECEPTOR FOR THE C-X-C CHEMOKINE CXCL16. USED AS A
 CC CORECEPTOR BY SIVS AND BY STRAINS OF HIV-2 AND M-TROPIC HIV-1.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN LYMPHOID TISSUES AND ACTIVATED T
 CC CELLS.
 CC
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 CC EMBL: AF007545; AAB64221.1; -
 DR EMBL: U73529; AAB61456.1; -
 DR EMBL: U73531; AAB61457.1; -
 DR EMBL: Y13248; CAA73698.1; -
 DR MIM: 605163; -
 DR InterPro: IPR000276; GPCR_Rhodpsn.
 DR Pfam: PF00001; 7tm_1; 1.


```

GN ADK OR SPY0074.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370 / ATCC 700294 / Serotype M1.
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., Moshan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezeau S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RN Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
RN [2]
RP SEQUENCE OF 1-13 AND 207-212, AND MASS SPECTROMETRY.
RC STRAIN-JR34 / Serotype M6;
RA Hogan D.A., Du P., Stevenson T.I., Whilton M., Kilby G.W., Rogers J.,
RA Vanbogaert R.A.;
RT "Two-dimensional gel electrophoresis map of Streptococcus pyogenes
RT proteins";
RL Submitted (MAY-2000) to the SWISS-PROT data bank.
CC -1- FUNCTION: THIS SMALL OBLIGATORY ENZYME IS ESSENTIAL FOR
CC MAINTENANCE AND CELL GROWTH.
CC -1- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- MASS SPECTROMETRY: MW=23699.98; METHOD=electrospray.
CC -1- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE006478; AAK33203.1; -
DR HSSP: P27142; 1ZTN.
DR InterPro: IPR000850; Adenylate_kin.
DR Pfam: PF00406; adenylatekinase; 1.
DR PRINTS: PR00094; ADENYLTKINASE.
DR PRODOM: PD000657; Adenylate_kin; 1.
DR PROSITE: PS00113; ADENYLATE_KINASE; 1.
DR KW Transferrase; kinase; ATP-binding; Complete proteome.
FT NP_BIND 7 15 ATP (BY SIMILARITY).
SQ SEQUENCE 212 AA; 23700 MW; F98AA65E98B39DEF CRC64;

```

```

Query Match 37.7%; Score 43; DB 1; Length 212;
Best Local Similarity 53.8%; Pred. No. 9.8;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 8 EDYGFAMITRADM 20
    1 1 1 1 1 1 1
Db 22 EEFGLAHISTGDM 34

```

```

RESULT 11.
YCAJ_COXBU STANDARD: PRT: 446 AA.
AC P39918;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 49.9 kDa protein in SpoIIIE-ers intergenic region.
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Coxiella group; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Oswald W.;
RL Thesis (1994), Justus Liebig University / Frankfurt, Germany.
RN [2]
RP CONCEPTUAL TRANSLATION.
RA Bairoch A.;
RL Unpublished observations (AUG-1995).
CC -1- SIMILARITY: STRONG, TO E.COLI YCAJ AND H.INFLUENZAЕ HOMOLOG.
CC -1- SIMILARITY: TO YEAST YNL218W.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION: A FRAMESHIFT WAS
CC INTRODUCED IN POSITION 66 TO PRODUCE THIS ORF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X75627; CAA53291.1; ALT_FRAME.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003959; AAA_ATPase-centr.
DR InterPro: IPR000862; RFDomain.
DR Pfam: PF00004; AAA; 1.
DR SMART: SM00382; AAA; 1.
DR KW Hypothetical protein; ATP-binding.
FT NP_BIND 51 58 ATP (POTENTIAL).
SQ SEQUENCE 446 AA; 49875 MW; D24877B3C6A4E04B CRC64;

```

```

Query Match 37.7%; Score 43; DB 1; Length 446;
Best Local Similarity 57.1%; Pred. No. 21;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 1 MNSGDYKEDYGFAN 14
    1 1 1 1 1 1 1
Db 377 MKSLDYKDYRYAH 390

```

```

RESULT 12
AMYG_DEBOC STANDARD: PRT: 958 AA.
AC P22861; Q92336;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Glucoamylase 1 precursor (FC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GAM1.
OS Debaryomyces occidentalis (Yeast) (Schwanniomyces occidentalis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX NCBI_TaxID=27300;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-ATCC 26076;
RX MEDLINE=91071592; PubMed=1979298;
RA Dohmen R.J., Strasser A.W.M., Dahlems U.M., Hollenberg C.P.;
RT "Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1)
RT and its expression in Saccharomyces cerevisiae.";
RL Gene 95:111-121(1990).
RN [2]
RP SIMILARITY TO OTHER FAMILY 31 MEMBERS.
RX MEDLINE=92077121; PubMed=1743281;
RA Naim H.Y., Niemann T., Kleinans U., Hollenberg C.P.,
RA Strasser A.W.M.;
RT "Striking structural and functional similarities suggest that
RT intestinal sucrose-isomaltase, human lysosomal alpha-glucosidase and
RT Schwanniomyces occidentalis glucoamylase are derived from a common
RT ancestral gene.";
RL FEBS Lett 294:109-112(1991).
CC -1- FUNCTION: THIS GLUCOAMYLASE HAS A SPECIFICITY TOWARD BOTH
CC ALPHA-1,4 AND ALPHA-1,6 LINKAGES.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-

```



```

CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M60207; AAA33923.1;
CC PIR: JN0102;
CC InterPro: IPR000322; Glyco_hydro_31.
CC Pfam: PF01055; Glyco_hydro_31.1.
CC PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
CC PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; 1.
CC K01001; Glycosylase; Polysaccharide degradation; glycoprotein;
CC signal.
CC FT SIGNAL 1 22 POTENTIAL.
CC FT CHAIN 23 958 GLUCAMYLASE 1.
CC FT ACT_SITE 470 470 BY SIMILARITY.
CC FT DOMAIN 26 41 SER-RICH.
CC FT DOMAIN 530 542 SER/THR-RICH.
CC FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 416 416 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 580 580 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 813 813 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 907 907 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 958 AA; 106507 MW; 47938DB9BC308260 CRC64;

Query Match 37.7%; Score 43; DB 1; Length 958;
Best Local Similarity 57.1%; Pred. No. 47;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 GDYKEDYGFANHT 16
   ||||| 1:1
Db 896 SGDYKADQPLANVT 909

RESULT 13
EX5B_MVCTU STANDARD; PRT; 1094 AA.
AC P66920;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5).
GN RECB OR RV0630C OR MT0658 OR MTCY20H10.11C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RA [1]
RA SEQUENCE FROM N.A.
RA STRAIN=H37Rv;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sutton J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the

```

```

RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,
RA Bisht W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR. IT CATALYZES THE
CC UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
CC STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
CC ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-
CC phosphooligonucleotides.
CC -1- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UVRD SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Z92772; CAB07119.1;
CC EMBL: AE006961; AAK44882.1;
CC DR HSE; P56285; 2PVR.
CC DR TIGR; MT0658;
CC DR TubercuList; RV0630C;
CC DR InterPro: IPR004586; RECB.
CC DR InterPro: IPR000212; UvrD-helicase.
CC Pfam: PF00580; UvrD-helicase; 1.
CC TIGRfam: TIGR00609; recb; 1.
CC KW Hydrolyase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
CC NP_BIND 21 28 ATP (POTENTIAL).
CC SQ SEQUENCE 1094 AA; 118721 MW; 31262D376875C201 CRC64;

Query Match 37.7%; Score 43; DB 1; Length 1094;
Best Local Similarity 53.3%; Pred. No. 54;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 GDYKEDYGFANHT 18
   ||||| 1:1
Db 996 GDTADYGFERLTA 1010

RESULT 14
PURA_HAETN STANDARD; PRT; 1297 AA.
ID PUR4_HAETN
AC P43847;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phosphoribosylformylglycinamide synthase (EC 6.3.5.3) (FGAM
DE synthase) (Formylglycinamide ribotide amidotransferase) (FGARAT).
GN PUR4 OR H10752.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RA [1]
RA SEQUENCE FROM N.A.
RA STRAIN=Rd / KW20 / ATCC 51907;

```

```

RX MEDLINE=95350630; PubMed=7542800;
RA Flaischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Krelavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shilley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-
CC ribosyl)glycanamide + L-glutamine + H(2)O = ADP + phosphate + 2-
CC (formamido)-N(1)-(5-phospho-D-ribo-5')acetic acid + L-glutamate.
CC -1- PATHWAY: De novo purine biosynthesis; fourth step.
CC -1- SUBUNIT: MONOMER (BT SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE FGAMS
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32759; AAC2411.1; ALT_INIT.
CC TIGR: HI0752.
CC InterPro: IPR000728; AIRS-related.
CC Pfam: PF00586; AIRS; 1.
CC Pfam: PF02769; AIRS_C; 2.
CC Putine biosynthesis; Ligase: ATP-binding; Glutamine amidotransferase;
CC Complete proteome.
CC NP_BIND 308 319 ATP (POTENTIAL).
CC GATASE (BY SIMILARITY).
CC AC SITE 138 138
CC SEQUENCE. 1297 AA; 142747 MW; DCD23A9DA666E93 CRC64;
OY 1 MNSGDYKEDYGEFAHTR 17
Db 457 WDSGSKEDLPASVOR 473

RESULT 15
COBB_ONCMY STANDARD: PRT; 587 AA.
AC Q90X85;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement component C8 beta chain precursor.
GN C8B.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RA Zarkadis I.K., Kazantzzi A., Sfyroera G., Lambiris J.D.;
RT "Cloning and characterization of eighth beta component of complement
RT in rainbow trout."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: C8 is a constituent of the MEMBRANE ATTACK COMPLEX. C8
CC BINDS TO THE C5B-7 COMPLEX, FORMING THE C5B-8 COMPLEX. C5-B8 BINDS

```

```

CC C9 AND ACTS AS A CATALYST IN THE POLYMERIZATION OF C9 (By
CC similarity).
CC -1- SUBUNIT: C8 IS COMPOSED OF THREE CHAINS: ALPHA, BETA AND GAMMA.
CC THE BETA CHAIN BINDS TO THE C8 ALPHA CHAIN AND TO THE C5B-C7
CC COMPLEX, PRESUMABLY TO C5B. IT IS ESSENTIAL TO THE INCORPORATION
CC OF C8 INTO THE C5B-C8 COMPLEX (By similarity).
CC -1- SIMILARITY: TO COMPLEMENT FACTORS C6, C7, C9, AND TO PERFORIN.
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF418597; AAL16647.1; -
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR002172; LDL_recept_A.
CC InterPro: IPR001862; MAC_perforin.
CC InterPro: IPR000884; TSP1.
CC Pfam: PF00057; ldl_recept_a; 1.
CC Pfam: PF01823; MACPF; 1.
CC SMART: SM00181; EGF; 1.
CC PROSITE: PS00022; EGF_1; 1.
CC PROSITE: PS01186; EGF_2; 1.
CC PROSITE: PS50068; LDLRA_2; 1.
CC PROSITE: PS00279; MAC_perforin; 1.
CC PROSITE: PS50092; TSP1; 2.
CC KW Complement pathway; Complement alternate pathway; Glycoprotein;
CC KW Signal; Plasma; Membrane attack complex; Cytolysis; EGF-like domain;
CC KW Repeat; Transmembrane.
CC FT SIGNAL 1 31 POTENTIAL.
CC FT PROPEP 32 44 BY SIMILARITY.
CC FT CHAIN 45 587 COMPLEMENT COMPONENT C8 BETA CHAIN.
CC FT DOMAIN 56 112 TSP TYPE-1.
CC FT DOMAIN 117 159 LDL-RECEPTOR CLASS A.
CC FT TRANSMEM 308 324 POTENTIAL.
CC FT TRANSMEM 329 348 POTENTIAL.
CC FT DOMAIN 495 531 EGF-LIKE.
CC FT DOMAIN 541 587 TSP TYPE-1 2.
CC FT DISULFD 118 129 BY SIMILARITY.
CC FT DISULFD 123 142 BY SIMILARITY.
CC FT DISULFD 374 399 BY SIMILARITY.
CC SEQUENCE 587 AA; 65540 MW; 6D80417659FC879E CRC64;
OY 2 NSGDYKE---DYGEFAHTRADM 20
Db 325 NYGEYRQIYRDYGYHTEATL 346

```

Search completed: June 16, 2003, 12:07:24
Job time : 3.33402 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:03:38 ; Search time 2.58197 Seconds
(without alignments)
781.893 Million cell updates/sec

Title: US-09-986-682B-1
Perfect score: 114
Sequence: 1 MNSGDYKEDYGFHITRADML 21

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	59.6	473	2	levanusucrase (EC 2
2	66	57.9	472	2	levanusucrase (EC 2
3	59	51.8	489	2	levanusucrase (Impo
4	55	48.2	646	2	hypothetical prote
5	48	42.1	310	2	UDP-n-acetylenolpy
6	48	42.1	378	2	hypothetical prote
7	48	42.1	478	2	6-phospho-beta-glu
8	47	41.2	289	2	hypothetical prote
9	47	41.2	567	2	lipoprotein (impor
10	46	40.4	217	2	adenylate kinase (
11	46	40.4	365	2	hypothetical prote
12	46	40.4	677	2	hypothetical prote
13	45	39.5	608	2	hypothetical prote
14	45	39.5	728	2	probable very-long
15	45	39.5	729	2	hypothetical prote
16	45	39.5	749	2	hypothetical prote
17	45	39.5	882	2	probable mutator-1
18	45	39.5	895	2	hypothetical prote
19	45	39.5	942	2	mutator-like trans
20	45	39.5	942	2	mutator-like trans
21	45	39.5	1071	2	mutator-like trans
22	45	39.5	1127	2	protein F27F5.15 (
23	44	38.6	250	2	probable mutator-1
24	44	38.6	364	2	pentapeptide repea
25	44	38.6	484	2	hypothetical prote
26	44	38.6	533	2	leucine aminopepti
27	43	37.7	227	2	RNA-directed RNA p
28	43	37.7	353	2	cytidylate kinase
29	43	37.7	457	2	hypothetical prote
					4-aminobutylate tr

30	43	37.7	478	2	6-phospho-beta-glu
31	43	37.7	958	1	glucan 1,4-alpha-g
32	43	37.7	1094	2	probable reeb prot
33	43	37.7	1320	2	phosphoribosylform
34	43	37.7	1948	2	N conserved hypoch
35	42	36.8	143	2	hypothetical prote
36	42	36.8	143	2	hypothetical prote
37	42	36.8	193	1	adenylate kinase (
38	42	36.8	198	2	NAD(P)H oxidoreduc
39	42	36.8	204	2	uridine-monophosph
40	42	36.8	220	2	adenylate kinase (
41	42	36.8	321	2	probable succinate
42	42	36.8	332	2	pyruvate formate-1
43	42	36.8	406	2	probable glycosyl
44	42	36.8	477	2	beta-glucosidase (
45	42	36.8	477	2	beta-glucosidase (

ALIGNMENTS

RESULT 1
A25040
levanusucrase (EC 2.4.1.10) sacB precursor - Bacillus subtilis
N:Alternate names: sucrose 6-fructosyl-transferase
C:Species: Bacillus subtilis
C:Date: 16-Aug-1988 #sequence_revision 20-Feb-1995 #text-change 20-Jun-2000
C:Accession: S07309; A25040; I39667; H69702
R:Steinmetz, M.; le Coq, D.; Aymerich, S.; Gonzy-Treboul, G.; Gay, P.
Mol. Gen. Genet. 200, 220-228, 1985
A:Title: The DNA sequence of the gene for the secreted Bacillus subtilis enzyme levan
A:Reference number: S07309; MUID:85295507; PMID:2993818
A:Accession: S07309
A:Molecule type: DNA
A:Residues: 1-473 <STE>
A:Cross-references: EMBL:X02730; NID:940118; PIDN:CAA26513.1; PID:g732568
A:Experimental source: Marburg
R:Shimotsu, H.; Henner, D.J.
J. Bacteriol. 168, 380-388, 1986
A:Title: Modulation of Bacillus subtilis levanusucrase gene expression by sucrose and
A:Reference number: A25040; MUID:87008406; PMID:2428811
A:Accession: A25040
A:Molecule type: DNA
A:Residues: 1-68 <SHD>
A:Cross-references: GB:M14202; NID:g143485; PIDN:AAA22725.1; PID:g143486
R:Fouet, A.; Arnaut, M.; Klier, A.; Rapoport, G.
Biochem. Biophys. Res. Commun. 119, 795-800, 1984
A:Title: Characterization of the precursor form of the exocellular levanusucrase from B
A:Reference number: I39667; MUID:84178454; PMID:6424671
A:Accession: I39667
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-11, '1', 13-62 <RES>
A:Cross-references: GB:K01987; NID:g143483; PIDN:AAA22724.1; PID:g143484
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Buschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabelt, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fuma, S.; Galizzi, A.; Gal
lechi, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
Y. M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstira, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipal, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zunsstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: H69702
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Title: Mechanisms of Evolution in *Rickettsia conorii* and *Rickettsia prowazekii*
A:Reference number: A97700; MOID:21442074; PMID:11557893
A:Accession number: D97741
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-378 <KUR>
A:Cross-References: GB:AE006914; PIDN: AAL02870.1; PID: g15619393; GSPDB: GN00173
C:Genetics:
A:Gene: murlb

Query Match 42.1%; Score 48; DB 2; Length 378;
Best Local Similarity 53.8%; Pred. No. 7;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 GQKEDYGFHIT 16
DB 92 GQKEDYGFHIT 104

RESULT 7
H82185
6-phospho-beta-glucosidase VC1558 [Imported] - *Vibrio cholerae* (strain N16961 serogroup C); Species: *Vibrio cholerae*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: H82185
R:Heidelberg, J.F.; Eissen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.; Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F. I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
N:ature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: H82185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-478 <HEI>
A:Cross-References: GB:AE004233; GB:AE003852; NID:99656055; PIDN:AAF94712.1; GSPDB: GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1558
A:Map position: 1
C:Superfamily: Agrobacterium beta-glucosidase

Query Match 42.1%; Score 48; DB 2; Length 478;
Best Local Similarity 41.2%; Pred. No. 9;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 3 SGDKEDYGFHITRAD 19
DB 434 TGEYKKRYGFIYVDKHD 450

RESULT 8
T05623
hypothetical protein F20D10.80 - *Arabidopsis thaliana*
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jun-1999
C:Accession: T05623
R:Bevan, M.; Weiler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F. submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15420
A:Accession: T05623
A:Molecule type: DNA
A:Residues: 1-289 <BEV>
A:Cross-References: EMBL:AL035538
A:Experimental source: cultivar Columbia; BAC clone F20D10
C:Genetics:
A:Map position: 4
A:Introns: 27/3; 75/1; 130/2; 169/2
A:Note: F20D10.80

Query Match 41.2%; Score 47; DB 2; Length 289;
Best Local Similarity 53.8%; Pred. No. 7.5;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 5 DYKEDYGFHITR 17
DB 157 DYKEDYGFHITR 169

RESULT 9
G90537
lipoprotein [Imported] - *Mycoplasma pulmonis* (strain UAB CTIP)
C:Species: *Mycoplasma pulmonis*
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: G90537
R:Chamblaud, I.; Heilig, R.; Parais, S.; Barbe, V.; Samson, D.; Galissou, F.; Moszer, Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen *Mycoplasma p*
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: G90537
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-567 <KUR>
A:Cross-References: GB:AL445566; PID: g14089620; PIDN: CAC13380.1; GSPDB: GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPV 2070
A:Genetic code: SGC3

Query Match 41.2%; Score 47; DB 2; Length 567;
Best Local Similarity 44.4%; Pred. No. 16;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 NSGDKEDYGFHITRAD 19
DB 435 HGGNKVDYFTHYNTN 452

RESULT 10
J50492
adenylate kinase (EC 2.7.4.3) - *Bacillus subtilis*
M:Alternate names: ATP-AMP transphosphorylase
C:Species: *Bacillus subtilis*
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Jun-2000
C:Accession: J50492; S12684; S08630; E69583
R:Nakamura, K.; Nakamura, A.; Takamatsu, H.; Yoshikawa, H.; Yamane, K. J. Biochem. 107, 603-607, 1990
A:Title: Cloning and characterization of a *Bacillus subtilis* gene homologous to E. co
A:Reference number: J50490; MUID:90292990; PMID:2113521
A:Accession: J50492
A:Molecule type: DNA
A:Residues: 1-217 <NAK>
A:Cross-References: GB:D00619; NID:9216336; PIDN: BAA00496.1; PID: 9216340
R:Toshikawa, H.; Doi, R.H. Nucleic Acids Res. 18, 1647, 1990
A:Title: Sequence of the *Bacillus subtilis* spectinomycin resistance gene region.
A:Reference number: S12680; MUID:90221911; PMID:2139212
A:Accession: S12684
A:Molecule type: DNA
A:Residues: 1-116 <ROS>
A:Cross-References: EMBL:M31102; NID:91184272; PIDN: AAB59119.1; PID: 9143579
A:Experimental source: strain 1A241
R:Sub, J.W.; Boylan, S.A.; Thomas, S.M.; Dolan, K.M.; Oliver, D.B.; Price, C.W. Mol. Microbiol. 4, 305-314, 1990
A:Title: Isolation of a secY homologue from *Bacillus subtilis*: evidence for a common
A:Reference number: S08628; MUID:90251170; PMID:2110998
A:Accession: S08630
A:Molecule type: DNA
A:Residues: 1-99 <SUH>
A:Cross-References: EMBL:X51329; NID:940132; PIDN: CAA35713.1; PID: 940135
A:Experimental source: strain Marburg; cell line PB2
R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; A.; Enrich, S.D.; Emerson, P.T.; Entlan, K.D.; Errington, J.; Fabret, C.; Ferrati, Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gal iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hultio Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Serot, A.; Schuch, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tonon, A.; Tosato, V.; Uchiyama, T.; Walters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; M0ID:98044033; PMID:9384377
 A:Accession: E69583
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1217 <KUN>
 A:Cross-references: GB:299104; GB:AL009126; NID:92632267; PIDN:CB11913.1; PID:92632404
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: ack
 C:Function:
 A:Description: catalyzes reversible phosphorylation of AMP with ATP to form two ADP
 A:Note: magnesium required
 C:Superfamily: adenylate kinase
 C:Keywords: ATP; P-loop; phosphotransferase
 F:7-14/Region: nucleotide-binding motif A (P-loop) #status atypical
 F:80-85/Region: nucleotide-binding motif B #status atypical
 F:28,84/Active site: His, Asp #status predicted

Query Match 40.4%; Score 46; DB 2; Length 217;
 Best Local Similarity 61.5%; Pred. No. 7.9;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 8 EDYGFHITRADM 20
 1111 11: 11
 Db 22 EDYGFHITGDM 34

RESULT 11
 A71048
 hypothetical protein PH1673 - *Pyrococcus horikoshii*
 C:Species: *Pyrococcus horikoshii*
 C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
 C:Accession: A71048
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Seki, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
 A:Reference number: A71000; M0ID:98344137; PMID:9679194
 A:Accession: A71048
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1365 <KAW>
 A:Cross-references: GB:AP00006; NID:93236133; PIDN:BA30785.1; PID:93258102
 A:Experimental source: strain OT3
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
 C:Genetics:
 A:Gene: PH1673
 C:Superfamily: *Pyrococcus abyssi* hypothetical protein PAB0337

Query Match 40.4%; Score 46; DB 2; Length 365;
 Best Local Similarity 47.1%; Pred. No. 14;
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 2 NSGDYKEDYGFHITRA 18
 1111 11: 11: 11
 Db 249 NSGDYKEDYGFHITRA 265

RESULT 12
 G84614
 hypothetical protein At2g22620 [imported] - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: G84614
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shua, T.P.; Benito, M.T.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
 A:Reference number: A84420; M0ID:20083487; PMID:1067197
 A:Accession: G84614
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-677 <STO>
 A:Cross-references: GB:AE002093; NID:94314359; PIDN:AAD15570.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g22620
 A:Map position: 2

Query Match 40.4%; Score 46; DB 2; Length 677;
 Best Local Similarity 53.8%; Pred. No. 28;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 5 DYKEDYGFHITR 17
 11: 11: 11: 1
 Db 545 DYKEDYGFHITR 557

RESULT 13
 H83284
 Probable very-long-chain acyl-CoA synthetase PA2893 [imported] - *Pseudomonas aeruginosa*
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: H83284
 R:Stover, C.K.; Pham, X.Q.; EWlin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L. Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa
 A:Reference number: A82950; M0ID:20437337; PMID:10984043
 A:Accession: H83284
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-608 <STO>
 A:Cross-references: GB:AE004715; GB:AE004091; NID:99948977; PIDN:AAG06281.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA2893
 C:Superfamily: *Mycobacterium tuberculosis* probable fadD6 protein; acetate-CoA ligase

Query Match 39.5%; Score 45; DB 2; Length 608;
 Best Local Similarity 45.0%; Pred. No. 36;
 Matches 9; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

OY 2 NSGDYKEDYGFHITRADML 21
 1111 11: 11: 11: 1
 Db 450 NSGDYKEDYGFHITRADML 469

RESULT 14
 D86233
 hypothetical protein [imported] - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: D86233
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon, Chin, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Hulzar, L. Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Matzla, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H. A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W. A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
 A:Reference number: A86141; M0ID:21016719; PMID:11130712
 A:Accession: D86233
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-728 <STO>

A:Cross-references: GB:AE005173; NID:92160180; PIDN:AAB60743.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 39.5%; Score 45; DB 2; Length 728;
 Best Local Similarity 53.8%; Pred. No. 44;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 5 DYKEDYGFHITR 17
 ||:|:||||:
 Db 597 DYRDWFFAHVPR 609

RESULT 15

G96559

hypothetical protein F5F19.7 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002

C:Accession: G96559

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzalli,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G96559

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-729 <STO>

A:Cross-references: GB:AE005173; NID:94220448; PIDN:AD12675.1; GSPDB:GN00141

C:Genetics:

A:Gene: F5F19.7

A:Map position: 1

C:Superfamily: Arabidopsis thaliana hypothetical protein F26C24.7

Query Match 39.5%; Score 45; DB 2; Length 729;
 Best Local Similarity 69.2%; Pred. No. 44;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 6 YKEDYGFHITRA 18
 |||||
 Db 663 YLEDIGFEHWTRA 675

Search completed: June 16, 2003, 12:10:40
 Job time : 4.58197 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:07:33 : Search time 3.22746 Seconds
(without alignments)
695.665 Million cell updates/sec

Title: US-09-986-682B-1
Perfect score: 114
Sequence: 1 MNSGDYKEDYGFHITRADML 21

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: Published_Applications_AA:*
2: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB pep:*
3: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB pep:*
5: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB pep:*
6: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB pep:*
7: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB pep:*
8: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB pep:*
9: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB pep:*
12: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB pep:*
14: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	100.0	21	10 US-09-986-682B-1	Sequence 1, Appli
2	114	100.0	455	10 US-09-986-682B-3	Sequence 3, Appli
3	48	42.1	310	9 US-09-991-211-8	Sequence 8, Appli
4	44	38.6	18	9 US-10-084-813-201	Sequence 201, App
5	44	38.6	18	9 US-10-084-813-424	Sequence 424, App
6	44	38.6	342	10 US-09-852-156-2	Sequence 2, Appli
7	44	38.6	342	10 US-09-940-063-2	Sequence 2, Appli
8	43	37.7	18	9 US-10-084-813-202	Sequence 202, App
9	43	37.7	18	9 US-10-084-813-425	Sequence 425, App
10	43	37.7	18	9 US-10-084-813-426	Sequence 426, App
11	43	37.7	18	9 US-10-084-813-427	Sequence 427, App
12	43	37.7	18	9 US-10-084-813-428	Sequence 428, App
13	42	36.8	143	9 US-09-928-457-39	Sequence 39, Appli
14	42	36.8	364	9 US-10-153-668-472	Sequence 472, App
15	42	36.8	366	9 US-10-153-668-472	Sequence 472, App
16	42	36.8	479	10 US-09-815-242-10288	Sequence 10288, A
17	42	36.8	566	9 US-09-797-500-2	Sequence 2, Appli
18	41	36.0	126	10 US-09-815-242-14024	Sequence 14024, A
19	41	36.0	242	9 US-09-764-868-1177	Sequence 1177, Ap

20	41	36.0	245	9 US-09-764-868-865	Sequence 865, App
21	41	36.0	386	9 US-09-764-868-866	Sequence 866, App
22	40.5	35.5	504	9 US-10-158-895-2	Sequence 2, Appli
23	40.5	35.5	504	12 US-10-123-427-2	Sequence 2, Appli
24	40.5	35.5	504	12 US-10-123-427-6	Sequence 6, Appli
25	40.5	35.5	513	9 US-10-158-895-43	Sequence 43, Appli
26	40.5	35.5	517	9 US-10-158-895-11	Sequence 11, Appli
27	40	35.1	88	9 US-09-738-656-6794	Sequence 6794, Ap
28	39	34.2	60	10 US-09-864-761-36935	Sequence 36935, A
29	39	34.2	60	10 US-09-864-761-48469	Sequence 48469, A
30	39	34.2	194	9 US-09-791-279-155	Sequence 155, App
31	39	34.2	284	10 US-09-815-242-13239	Sequence 13239, A
32	39	34.2	477	10 US-09-966-147-4	Sequence 4, Appli
33	39	34.2	684	10 US-09-823-240-9	Sequence 9, Appli
34	39	34.2	822	10 US-09-966-147-2	Sequence 2, Appli
35	39	34.2	847	10 US-09-924-859A-5	Sequence 5, Appli
36	39	34.2	1016	10 US-09-815-242-5845	Sequence 5845, Ap
37	38.5	33.8	904	9 US-10-121-988-18	Sequence 18, Appli
38	38.5	33.8	904	10 US-09-894-998-18	Sequence 18, Appli
39	38	33.3	25	10 US-09-839-666-19	Sequence 19, Appli
40	38	33.3	26	10 US-09-839-666-15	Sequence 15, Appli
41	38	33.3	130	10 US-09-815-242-5156	Sequence 5156, Ap
42	38	33.3	193	10 US-09-925-300-1057	Sequence 1057, Ap
43	38	33.3	252	10 US-09-925-301-1273	Sequence 1273, Ap
44	38	33.3	478	10 US-09-815-242-5719	Sequence 5719, Ap
45	38	33.3	478	10 US-09-815-242-12480	Sequence 12480, A

ALIGNMENTS

RESULT 1
US-09-986-682B-1
Sequence 1, Application US/09986682B
Patent No. US20020115182A1
GENERAL INFORMATION:
APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU
KENKYUJO
TITLE OF INVENTION: POLYPEPTIDE HAVING --FRUCTOPURANOSIDASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/986,682B
FILING DATE: 06-NO. US20020115182A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI=2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
FRAGMENT TYPE: N-terminal fragment

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-986-682B-1

Query Match
Best Local Similarity 100.0%; Score 114; DB 10; Length 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFAHITRADML 21
Db 1 MNSGDYKEDYGFAHITRADML 21

RESULT 2
US-09-986-682B-3
Sequence 3, Application US/09986682B
Patent No. US20020115182A1
GENERAL INFORMATION:
APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU
KENKYUO.
TITLE OF INVENTION: POLYPEPTIDE HAVING --FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/986,682B
FILING DATE: 06-NO. US20020115182A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI=2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-986-682B-3

Query Match
Best Local Similarity 100.0%; Score 114; DB 10; Length 455;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFAHITRADML 21
Db 1 MNSGDYKEDYGFAHITRADML 21

RESULT 3
US-09-991-211-8
Sequence 8, Application US/09991211
Patent No. US20020156585A1
GENERAL INFORMATION:
APPLICANT: Pharmacia & Upjohn
TITLE OF INVENTION: CRYSTALLIZATION AND STRUCTURE DETERMINATION OF

TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS
TITLE OF INVENTION: UDP-N-ACETYLENOLPYRUVYLGLUCOSAMINE REDUCTASE (S. aureus)
FILE REFERENCE: 268 6241 0101
CURRENT APPLICATION NUMBER: US/09/991,211
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 09/632,947
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/147,164
PRIOR FILING DATE: 1999-08-04
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 8
LENGTH: 310
TYPE: PRT
ORGANISM: Rickettsia prowazekii
US-09-991-211-8

Query Match
Best Local Similarity 42.1%; Score 48; DB 9; Length 310;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 4 GDYKEDYGFAHIT 16
Db 24 GEYKDYNLKHLT 36

RESULT 4
US-10-084-813-201
Sequence 201, Application US/10084813
Publication No. US20030068615A1
GENERAL INFORMATION:
APPLICANT: SAXINGER, CARL
TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE
FILE REFERENCE: 215875
CURRENT APPLICATION NUMBER: US/10/084,813
CURRENT FILING DATE: 2002-02-27
PRIOR APPLICATION NUMBER: PCT/US00/23505
PRIOR FILING DATE: 2000-08-25
PRIOR APPLICATION NUMBER: US 60/151,270
PRIOR FILING DATE: 1999-08-27
NUMBER OF SEQ ID NOS: 1242
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 201
LENGTH: 18
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: binding peptide
US-10-084-813-201

Query Match
Best Local Similarity 38.6%; Score 44; DB 9; Length 18;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFA 13
Db 1 MAEDHYHEDYGS 13

RESULT 5
US-10-084-813-424
Sequence 424, Application US/10084813
Publication No. US20030068615A1
GENERAL INFORMATION:
APPLICANT: SAXINGER, CARL
TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE
FILE REFERENCE: 215875
CURRENT APPLICATION NUMBER: US/10/084,813
CURRENT FILING DATE: 2002-02-27
PRIOR APPLICATION NUMBER: PCT/US00/23505

PRIOR FILING DATE: 2000-08-25
PRIOR APPLICATION NUMBER: US 60/151,270
PRIOR FILING DATE: 1999-08-27
NUMBER OF SEQ ID NOS: 1242
SOFTWARE: PatentIn version 3.1
SEQ ID NO 424
LENGTH: 18
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: binding peptide
US-10-084-813-424

Query Match 38.6%; Score 44; DB 9; Length 18;
Best Local Similarity 61.5%; Pred. No. 1.3;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFA 13
1 111111:
Db 1 MAEHDYHEDYGFS 13

RESULT 6
US-09-852-156-2
Sequence 2, Application US/09852156
Patent No. US20020076694A1
GENERAL INFORMATION:
APPLICANT: Littmap, Dan R.
Deng, Hongkui
Unutmaz, Derya
Ramani, Vineet N.K.

TITLE OF INVENTION: NOVEL ALTERNATIVE G-COUPLED RECEPTORS
ASSOCIATED WITH RETROVIRAL ENTRY INTO CELLS, METHODS OF
IDENTIFYING THE SAME, AND DIAGNOSTIC AND THERAPEUTIC USES
THEREOF

NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:

ADDRESSEE: David A. Jackson, Esq.

STREET: 411 Hackensack Ave, Continental Plaza, 4th

Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/852,156

FILING DATE: 09-May-2001

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 1049-1-009NCIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 342 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Query Match 38.6%; Score 44; DB 10; Length 342;
Best Local Similarity 61.5%; Pred. No. 3.1;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFA 13
1 111111:
Db 1 MAEHDYHEDYGFS 13

RESULT 7
US-09-940-063-2
Sequence 2, Application US/09940063
Patent No. US20020090657A1
GENERAL INFORMATION:
APPLICANT: Briskin, Michael J.
APPLICANT: Murphy, Kristine E.
APPLICANT: Wilbanks, Alyson M.
APPLICANT: Wu, Lijun
TITLE OF INVENTION: No. US20020090657A1el Antibodies and Ligands for "bonzo"
FILE REFERENCE: 1855,1070-000
CURRENT APPLICATION NUMBER: US/09/940,063
CURRENT FILING DATE: 2001-08-27
PRIOR APPLICATION NUMBER: 09/449,437
PRIOR FILING DATE: 1999-11-24
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 342
TYPE: PRT
ORGANISM: Homo sapiens
US-09-940-063-2

Query Match 38.6%; Score 44; DB 10; Length 342;
Best Local Similarity 61.5%; Pred. No. 3.1;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFA 13
1 111111:
Db 1 MAEHDYHEDYGFS 13

RESULT 8
US-10-084-813-202
Sequence 202, Application US/10084813
Publication No. US20030068615A1
GENERAL INFORMATION:
APPLICANT: SAXINGER, CARL
TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE
FILE REFERENCE: 215875
CURRENT APPLICATION NUMBER: US/10/084,813
CURRENT FILING DATE: 2002-02-27
PRIOR APPLICATION NUMBER: PCT/US00/23505
PRIOR FILING DATE: 2000-08-25
PRIOR APPLICATION NUMBER: US: 60/151,270
PRIOR FILING DATE: 1999-08-27
NUMBER OF SEQ ID NOS: 1242
SOFTWARE: PatentIn version 3.1
SEQ ID NO 202
LENGTH: 18
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: binding peptide
US-10-084-813-202

Query Match 37.7%; Score 43; DB 9; Length 18;
Best Local Similarity 77.8%; Pred. No. 1.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 5 DYKEDYGFA 13
1111111:

Db 1 DYHEDYGFS 9

RESULT 9

US-10-084-813-425
Sequence 425, Application US/10084813
Publication No. US20030068615A1

GENERAL INFORMATION:

APPLICANT: SAXINGER, CARL

TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE

FILE REFERENCE: 215875

CURRENT APPLICATION NUMBER: US/10/084,813

PRIOR FILING DATE: 2002-02-27

PRIOR APPLICATION NUMBER: PCT/US00/23505

PRIOR FILING DATE: 2000-08-25

PRIOR APPLICATION NUMBER: US 60/151,270

PRIOR FILING DATE: 1999-08-27

NUMBER OF SEQ ID NOS: 1242

SOFTWARE: PatentIn version 3.1

SEQ ID NO 425

LENGTH: 18

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: binding peptide

US-10-084-813-425

Query Match
Best Local Similarity 37.7%; Score 43; DB 9; Length 18;
Best Local Similarity 77.8%; Pred. No. 1.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 DYKEDYGFA 13
11111111

Db 4 DYHEDYGFS 12

RESULT 10

US-10-084-813-426

Sequence 426, Application US/10084813
Publication No. US20030068615A1

GENERAL INFORMATION:

APPLICANT: SAXINGER, CARL

TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE

FILE REFERENCE: 215875

CURRENT APPLICATION NUMBER: US/10/084,813

PRIOR FILING DATE: 2002-02-27

PRIOR APPLICATION NUMBER: PCT/US00/23505

PRIOR FILING DATE: 2000-08-25

PRIOR APPLICATION NUMBER: US 60/151,270

PRIOR FILING DATE: 1999-08-27

NUMBER OF SEQ ID NOS: 1242

SOFTWARE: PatentIn version 3.1

SEQ ID NO 426

LENGTH: 18

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: binding peptide

US-10-084-813-426

Query Match
Best Local Similarity 37.7%; Score 43; DB 9; Length 18;
Best Local Similarity 77.8%; Pred. No. 1.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 DYKEDYGFA 13
11111111

Db 3 DYHEDYGFS 11

RESULT 11

US-10-084-813-427

Sequence 427, Application US/10084813
Publication No. US20030068615A1

GENERAL INFORMATION:

APPLICANT: SAXINGER, CARL

TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE

FILE REFERENCE: 215875

CURRENT APPLICATION NUMBER: US/10/084,813

PRIOR FILING DATE: 2002-02-27

PRIOR APPLICATION NUMBER: PCT/US00/23505

PRIOR FILING DATE: 2000-08-25

PRIOR APPLICATION NUMBER: US 60/151,270

PRIOR FILING DATE: 1999-08-27

NUMBER OF SEQ ID NOS: 1242

SOFTWARE: PatentIn version 3.1

SEQ ID NO 427

LENGTH: 18

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: binding peptide

US-10-084-813-427

Query Match
Best Local Similarity 37.7%; Score 43; DB 9; Length 18;
Best Local Similarity 77.8%; Pred. No. 1.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 DYKEDYGFA 13
11111111

Db 2 DYHEDYGFS 10

RESULT 12

US-10-084-813-428

Sequence 428, Application US/10084813
Publication No. US20030068615A1

GENERAL INFORMATION:

APPLICANT: SAXINGER, CARL

TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE

FILE REFERENCE: 215875

CURRENT APPLICATION NUMBER: US/10/084,813

PRIOR FILING DATE: 2002-02-27

PRIOR APPLICATION NUMBER: PCT/US00/23505

PRIOR FILING DATE: 2000-08-25

PRIOR APPLICATION NUMBER: US 60/151,270

PRIOR FILING DATE: 1999-08-27

NUMBER OF SEQ ID NOS: 1242

SOFTWARE: PatentIn version 3.1

SEQ ID NO 428

LENGTH: 18

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: binding peptide

US-10-084-813-428

Query Match
Best Local Similarity 37.7%; Score 43; DB 9; Length 18;
Best Local Similarity 77.8%; Pred. No. 1.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 DYKEDYGFA 13
11111111

Db 1 DYHEDYGFS 9

RESULT 13

US-09-928-457-39

Sequence 39, Application US/09928457
Patent No. US20020164603A1

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: DNA, specific proteins and peptides

TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method
TITLE OF INVENTION: for obtaining them and their biological application.
NUMBER OF SEQUENCES: 99
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30 (OEB)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/928,457
FILING DATE: 2001-08-14
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/214,759
FILING DATE: 199-12-10
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 143 acids amin.s
TYPE: acide amin,
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..143
US-09-928-457-39

Query Match 36.8%; Score 42; DB 9; Length 143;
Best Local Similarity 53.3%; Pred. No. 26;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 4 GDYKEDYGFAHITRA 18
111:111:11
Db 67 GDYEEVNGFEYIDKA 81

RESULT 14
US-10-153-668-472
Sequence 472, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:
APPLICANT: HONDA, Goichi
APPLICANT: MATSUDA, Akio
APPLICANT: MURAMATSU, Shuji
APPLICANT: ISHIZAWA, Kenya
TITLE OF INVENTION: STAT6 Activating Gene
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
PRIOR FILING DATE: 2001-10-10
NUMBER OF SEQ ID NOS: 488
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 472
LENGTH: 364
TYPE: PRT
ORGANISM: Homo sapiens
US-10-153-668-472

Query Match 36.8%; Score 42; DB 9; Length 364;
Best Local Similarity 46.7%; Pred. No. 69;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 5 DYKEDYGFAHITRAD 19

Db 108 DIVKDYAFVHMERAE 122

RESULT 15
US-10-153-668-228
Sequence 228, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:
APPLICANT: HONDA, Goichi
APPLICANT: MATSUDA, Akio
APPLICANT: MURAMATSU, Shuji
APPLICANT: ISHIZAWA, Kenya
TITLE OF INVENTION: STAT6 Activating Gene
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
PRIOR FILING DATE: 2001-10-10
NUMBER OF SEQ ID NOS: 488
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 228
LENGTH: 366
TYPE: PRT
ORGANISM: Homo sapiens
US-10-153-668-228

Query Match 36.8%; Score 42; DB 9; Length 366;
Best Local Similarity 46.7%; Pred. No. 69;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 5 DYKEDYGFAHITRAD 19
111:111:11
Db 108 DIVKDYAFVHMERAE 122

Search completed: June 16, 2003, 12:22:35
Job time : 5.22746 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:04:38 ; Search time 1.89344 Seconds
(without alignments)
326.327 Million cell updates/sec

Title: US-09-986-682B-1

Perfect score: 114

Sequence: 1 MNSGDYKEDYGFATRADML 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	100.0	21	2	US-08-870-827-1
2	114	100.0	21	2	US-09-317-179-1
3	114	100.0	455	2	US-08-870-827-3
4	114	100.0	455	4	US-09-317-179-3
5	48	42.1	310	4	US-09-632-947B-8
6	45.5	39.9	133	4	US-09-134-001C-4596
7	44	38.6	342	2	US-08-742-011-2
8	44	38.6	342	4	US-09-275-384B-5
9	44	38.6	342	4	US-09-116-498-2
10	44	38.6	342	4	US-09-449-437A-2
11	44	38.6	342	4	US-09-517-605-9
12	43	37.7	425	4	US-09-134-001C-4433
13	42	36.8	566	1	US-08-073-383-4
14	42	36.8	566	1	US-08-428-415-4
15	42	36.8	566	1	US-08-379-685-4
16	42	36.8	566	3	US-08-854-029-4
17	42	36.8	566	3	US-08-848-810-2
18	42	36.8	566	4	US-08-428-762-4
19	42	36.8	566	5	PCT-US94-06365-4
20	41	36.0	197	2	US-08-879-561-5
21	40.5	35.5	504	2	US-08-752-891-2
22	40.5	35.5	504	2	US-08-752-891-6
23	40.5	35.5	504	2	US-09-144-178-2
24	40.5	35.5	504	2	US-09-144-178-6
25	40.5	35.5	504	4	US-09-406-854-2
26	40.5	35.5	504	4	US-09-406-854-6
27	40.5	35.5	504	4	US-09-529-279-2

28	40.5	35.5	513	4	US-09-529-279-43	Sequence 43, Appl
29	40.5	35.5	517	4	US-09-529-279-11	Sequence 11, Appl
30	39.5	34.6	891	1	US-08-042-747A-6	Sequence 6, Appl
31	39	34.2	194	2	US-08-879-561-11	Sequence 11, Appl
32	39	34.2	477	2	US-08-359-705B-4	Sequence 4, Appl
33	39	34.2	477	2	US-08-286-846A-4	Sequence 4, Appl
34	39	34.2	477	2	US-08-457-880A-4	Sequence 4, Appl
35	39	34.2	477	3	US-08-444-622A-4	Sequence 4, Appl
36	39	34.2	477	3	US-08-942-562-4	Sequence 4, Appl
37	39	34.2	477	3	US-09-156-923-4	Sequence 4, Appl
38	39	34.2	821	1	US-08-339-578-2	Sequence 2, Appl
39	39	34.2	822	2	US-08-359-705B-2	Sequence 2, Appl
40	39	34.2	822	2	US-08-286-846A-2	Sequence 2, Appl
41	39	34.2	822	2	US-08-457-880A-2	Sequence 2, Appl
42	39	34.2	822	3	US-08-444-622A-2	Sequence 2, Appl
43	39	34.2	822	3	US-08-942-562-2	Sequence 2, Appl
44	39	34.2	822	3	US-09-156-923-2	Sequence 2, Appl
45	39	34.2	847	1	US-08-286-305A-5	Sequence 5, Appl

ALIGNMENTS

```
RESULT 1
US-08-870-827-1
; Sequence 1, Application US/08870827
; Patent No. 5962297
; GENERAL INFORMATION:
; APPLICANT: Tsusaki et al.
; TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: BROWDY AND NEWMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,827
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 170,630/1996
; FILING DATE: 10-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI=2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; FRAGMENT TYPE: N-terminal fragment
; US-08-870-827-1

Query Match 100.0%; Score 114; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.4e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 MNSGDYKEDYGFATRADML 21
|||||
DB 1 MNSGDYKEDYGFATRADML 21

RESULT 2
US-09-317-179-1
Sequence 1, Application US/09317179
Patent No. 6383769
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/317,179
FILING DATE: 24-May-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/870,827
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
FRAGMENT TYPE: N-terminal fragment
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-317-179-1
Query Match 100.0%; Score 114; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.4e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNSGDYKEDYGFHITRADML 21
Db 1 MNSGDYKEDYGFHITRADML 21
RESULT 3
US-08-870-827-3
Sequence 3, Application US/08870827
Patent No. 5962297
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,827
FILING DATE: 06-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-870-827-3
Query Match 100.0%; Score 114; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNSGDYKEDYGFHITRADML 21
Db 1 MNSGDYKEDYGFHITRADML 21
RESULT 4
US-09-317-179-3
Sequence 3, Application US/09317179
Patent No. 6383769
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/317,179
FILING DATE: 24-May-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/870,827
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-317-179-3

Query Match 100.0%; Score 114; DB 4; Length 455;
Best Local Similarity 100.0%; Pred. No. 1,1e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MNSGDYKEDYGFATRADML 21
Db 1 MNSGDYKEDYGFATRADML 21

RESULT 5
US-09-632-947B-8

Sequence 8, Application US/09632947B
Patent No. 6356845

GENERAL INFORMATION:

APPLICANT: Pharmacia & Upjohn

TITLE OF INVENTION: CRYSTALLIZATION AND STRUCTURE DETERMINATION OF

TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS

TITLE OF INVENTION: UDP-N-ACETYLENOLPYRUVYLGLUCOSAMINE REDUCTASE (S. aureus)

FILE REFERENCE: 268.6241 0101

CURRENT FILING DATE: 2000-08-04

PRIOR APPLICATION NUMBER: 60/147,164

PRIOR FILING DATE: 1999-08-04

NUMBER OF SEQ ID NOS: 11

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 8

LENGTH: 310

TYPE: PRT

ORGANISM: Rickettsia prowazekii

US-09-632-947B-8

Query Match 42.1%; Score 48; DB 4; Length 310;
Best Local Similarity 53.8%; Pred. No. 3.5;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 4 GDYKEDYGFATHT 16
Db 24 GEYKKDYKHLHT 36

RESULT 6
US-09-134-001C-4596

Sequence 4596, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4596

LENGTH: 133

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4596

Query Match 39.9%; Score 45.5; DB 4; Length 133;
Best Local Similarity 34.5%; Pred. No. 3.4;
Matches 10; Conservative 5; Mismatches 5; Indels 9; Gaps 1;

Oy 1 MNSGD-----YKEDYGFATHTRADW 20

Db 54 LNDGSETPNVEIKYKDKKSHLEKADL 82

RESULT 7
US-08-742-011-2

Sequence 2, Application US/08742011

Patent No. 5824504

GENERAL INFORMATION:

APPLICANT: Elishourbagy, Nabil A.

APPLICANT: Bergsma, Derek J.

APPLICANT: Ellis, Catherine E.

TITLE OF INVENTION: Human 7-Transmembrane Receptor

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19406-2799

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/742,011

FILING DATE:

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Han, William T.

REGISTRATION NUMBER: 34,344

REFERENCE/DOCKET NUMBER: ATG50020P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5090

TELEFAX: 610-270-5219

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 342 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-742-011-2

Query Match 38.6%; Score 44; DB 2; Length 342;
Best Local Similarity 61.5%; Pred. No. 18;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 1 MNSGDYKEDYGF 13
Db 1 MAEHDYHEDYGF 13

RESULT 8
US-09-275-384B-5

Sequence 5, Application US/09275384B

Patent No. 6232084

GENERAL INFORMATION:

APPLICANT: MACPHER, COLIN HOUSTON

APPLICANT: MOORES, KITTY

TITLE OF INVENTION: NEW USE

FILE REFERENCE: GH-31106

CURRENT APPLICATION NUMBER: US/09/275,384B

CURRENT FILING DATE: 1999-03-24

PRIOR APPLICATION NUMBER: 9806677.2

PRIOR FILING DATE: 1998-03-27

NUMBER OF SEQ ID NOS: 9

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 5

LENGTH: 342

TYPE: PRT

ORGANISM: HOMO SAPIENS

US-09-275-384B-5

Query Match 38.6%; Score 44; DB 4; Length 342;

Best Local Similarity 61.5%; Pred. No. 18;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNSGDYKEDYGFA 13
1 111111
Db 1 MAEHYHEDYGFS 13RESULT 9
US-09-116-498-2
Sequence 2, Application US/09116498

Patent No. 6251582

GENERAL INFORMATION:

APPLICANT: Littman, Dan R.

Deng, Hongkui

Unulmaz, Derya

Ramani, Vineet N.R.

TITLE OF INVENTION: NOVEL ALTERNATIVE G-COUPLED RECEPTORS

ASSOCIATED WITH RETROVIRAL ENTRY INTO CELLS, METHODS OF

IDENTIFYING THE SAME, AND DIAGNOSTIC AND THERAPEUTIC USES

THEREOF

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESS: David A. Jackson, Esq.

STREET: 411 Hackensack Ave, Continental Plaza, 4th

FLOOR

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/116,498

FILING DATE: 16-Jul-1998

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 1049-1-009

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-3800

TELEFAX: 201-343-1684

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 342 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-116-498-2

Query Match 38.6%; Score 44; DB 4; Length 342;

Best Local Similarity 61.5%; Pred. No. 18;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNSGDYKEDYGFA 13
1 111111
Db 1 MAEHYHEDYGFS 13RESULT 10
US-09-449-437A-2

Sequence 2, Application US/09449437A

Patent No. 6319675

GENERAL INFORMATION:

APPLICANT: Briskin, Michael J.

APPLICANT: Murphy, Kristine E.

APPLICANT: Wilbanks, Alyson M.

APPLICANT: Wu, Lijun

TITLE OF INVENTION: No. 6319675el Antibodies and Ligands for "Bonzo"

TITLE OF INVENTION: Chemokine Receptor

FILE REFERENCE: 1855.1070-000

CURRENT APPLICATION NUMBER: US/09/449,437A

CURRENT FILING DATE: 2001-01-09

NUMBER OF SEQ ID NOS: 18

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 342

TYPE: PRT

ORGANISM: Homo sapiens

US-09-449-437A-2

Query Match 38.6%; Score 44; DB 4; Length 342;

Best Local Similarity 61.5%; Pred. No. 18;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNSGDYKEDYGFA 13
1 111111
Db 1 MAEHYHEDYGFS 13RESULT 11
US-09-517-605-9
Sequence 9, Application US/09517605

Patent No. 6391567

GENERAL INFORMATION:

APPLICANT: Littman, Dan R.

APPLICANT: Kwon, Douglas S.

APPLICANT: van Kooyk, Yvette

APPLICANT: Geljtenbeck, Theo

TITLE OF INVENTION: METHODS OF USING A FACILITATOR OF RETROVIRAL ENTRY INTO

TITLE OF INVENTION: CELLS

FILE REFERENCE: 1049-1-017

CURRENT APPLICATION NUMBER: US/09/517,605

CURRENT FILING DATE: 2000-03-02

NUMBER OF SEQ ID NOS: 17

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 9

LENGTH: 342

TYPE: PRT

ORGANISM: Homo sapiens

US-09-517-605-9

Query Match 38.6%; Score 44; DB 4; Length 342;

Best Local Similarity 61.5%; Pred. No. 18;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNSGDYKEDYGFA 13
1 111111
Db 1 MAEHYHEDYGFS 13RESULT 12
US-09-134-001C-4433
Sequence 4433, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO: 4433
LENGTH: 425
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4433

Query Match 37.7% Score 43; DB 4; Length 425;
Best Local Similarity 61.5%; Pred. No. 33;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 9 DYGFHITRADML 21
| | | | | | | | | |
DB 5 DKGFHITRADML 17

RESULT 13
US-08-073-383-4
Sequence 4, Application US/08073383
Patent No. 5443962
GENERAL INFORMATION:
APPLICANT: Draetta, Giulio
APPLICANT: Cottarel, Guillaume
APPLICANT: Damaguez, Veronique
TITLE OF INVENTION: [fillin "Insert Title of Application" (ASSAY AND REAGENTS FOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESS: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/073,383
FILING DATE: 19930604
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIT-009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 566 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-073-383-4

Query Match 36.8% Score 42; DB 1; Length 566;
Best Local Similarity 47.1%; Pred. No. 66;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 GDYKEDYGFHITRADML 20
| | | | | | | | | |
DB 249 GDTEDDGFVILSDL 265

RESULT 14

US-08-428-415-4
Sequence 4, Application US/08428415
Patent No. 5756335
GENERAL INFORMATION:
APPLICANT: Cold Spring Harbor Laboratory
TITLE OF INVENTION: No. 5756335el Human cdc25 Genes, Encoded
TITLE OF INVENTION: Products and Uses Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESS: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,415
FILING DATE: 24 April 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIT-019CP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7400
TELEFAX: 617-227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 566 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-428-415-4

Query Match 36.8% Score 42; DB 1; Length 566;
Best Local Similarity 47.1%; Pred. No. 66;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 GDYKEDYGFHITRADML 20
| | | | | | | | | |
DB 249 GDTEDDGFVILSDL 265

RESULT 15
US-08-379-685-4
Sequence 4, Application US/08379685
Patent No. 5770423
GENERAL INFORMATION:
APPLICANT: Cold Spring Harbor Laboratory
TITLE OF INVENTION: No. 5770423el Human cdc25 Genes, Encoded
TITLE OF INVENTION: Products and Uses Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESS: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/379,685
FILING DATE: 26 January 1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MI-019-DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7400
TELEFAX: 617-227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 566 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-379-685-4

Query Match 36.8%; Score 42; DB 1; Length 566;
Best Local Similarity 47.1%; Pred. No. 66;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Oy 4 GDYKEDYGFHITRADM 20
||: || || ||: ||:
Db 249 GDTEDDGFVDLLES DL 265

Search completed: June 16, 2003, 12:11:31
Job time : 2.89344 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:55:48 ; Search time 5.59426 seconds
(without alignments)
500.202 Million cell updates/sec

Title: US-09-986-682B-1

Perfect score: 114

Sequence: 1 MNSGDYKEDYGFHITRADML 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	100.0	21	19	AAW48307
2	114	100.0	487	19	AAW48306
3	68	59.6	62	11	AAW48300
4	66	57.9	137	22	ABG04300
5	66	57.9	165	22	ABG04294
6	66	57.9	175	22	ABG04304
7	66	57.9	194	22	ABG26186
8	66	57.9	231	22	ABG04299
9	66	57.9	241	22	ABG21816
10	66	57.9	242	22	ABG04290

11	66	57.9	256	22	ABG04297	Novel human diagno
12	66	57.9	266	22	AAU30761	Novel human secret
13	66	57.9	267	22	ABG26189	Novel human diagno
14	66	57.9	268	22	ABG10291	Novel human diagno
15	66	57.9	276	22	ABG04324	Novel human diagno
16	66	57.9	280	22	ABG14490	Novel human diagno
17	66	57.9	282	22	ABG24525	Novel human diagno
18	66	57.9	287	22	ABG21864	Novel human diagno
19	66	57.9	288	22	ABG04293	Novel human diagno
20	66	57.9	293	22	ABG05672	Novel human diagno
21	66	57.9	306	22	ABG04301	Novel human diagno
22	66	57.9	314	22	ABG04303	Novel human diagno
23	66	57.9	314	22	ABG10673	Novel human diagno
24	66	57.9	315	22	ABG26187	Novel human diagno
25	66	57.9	324	22	ABG10297	Novel human diagno
26	66	57.9	331	22	ABG04296	Novel human diagno
27	66	57.9	338	22	ABG21839	Novel human diagno
28	66	57.9	342	22	ABG04302	Novel human diagno
29	66	57.9	342	22	ABG10300	Novel human diagno
30	66	57.9	343	22	ABG04315	Novel human diagno
31	66	57.9	343	22	ABG26188	Novel human diagno
32	66	57.9	347	22	ABG18055	Novel human diagno
33	66	57.9	347	22	ABG26458	Novel human diagno
34	66	57.9	350	22	ABG04291	Novel human diagno
35	66	57.9	350	22	ABG13458	Novel human diagno
36	66	57.9	350	22	ABG21691	Novel human diagno
37	66	57.9	357	22	ABG10299	Novel human diagno
38	66	57.9	360	22	ABG10298	Novel human diagno
39	66	57.9	361	22	ABG26191	Novel human diagno
40	66	57.9	369	22	ABG21840	Novel human diagno
41	66	57.9	374	22	ABG04285	Novel human diagno
42	66	57.9	376	22	ABG10307	Novel human diagno
43	66	57.9	378	22	ABG07324	Novel human diagno
44	66	57.9	378	22	ABG21624	Novel human diagno
45	66	57.9	384	22	ABG10303	Novel human diagno

ALIGNMENTS

RESULT 1

AAW48307

AAW48307 standard; Protein; 21 AA.

AC AAW48307;

DT 04-JUN-1998 (first entry)

DE Bacillus sp. V230 beta-fructofuranosidase N-terminal fragment.

KW Beta-fructofuranosidase; xylosyl fructoside; isomaltosyl fructoside;

KW lactosucrose; fructosyltrehalose; anticariogenic; growth promoter;

KW Bifidobacteria; mineral absorption promoting activity; taste; texture;

KW cosmetic; pharmaceutical; sweetener.

XX

OS Bacillus sp.

XX

PN EP812915-A2.

XX

PD 17-DEC-1997.

XX

PF 10-JUN-1997; 97EP-0304032.

XX

PR 10-JUN-1996; 96JP-0170630.

XX

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

XX

PI Chaen H. Kubota M. Tsusaki K.

XX

DR WPI; 1998-034976/04.

XX

PT Bacillus beta-fructofuranosidase enzyme - useful for producing

fructo:furanosylated saccharide(s) or alcohol(s)

Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 2 NSGDYKEDYGFAMITRADML 21
| ||| || : |||| |||
Db 35 NQAKKETYGVSHITRDML 54

RESULT 6
ABG04304
ID ABG04304 standard; Protein: 175 AA.

AC ABG04304;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #4295.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS68491.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 20; SEQ ID NO 34663; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 175 AA;

Query Match 57.9%; Score 66; DB 22; Length 175;
Best Local Similarity 65.0%; Pred. No. 0.012;
Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 2 NSGDYKEDYGFAMITRADML 21
| ||| || : |||| |||
Db 35 NQAKKETYGVSHITRDML 54

RESULT 7
ABG26186
ID ABG26186 standard; Protein: 194 AA.

AC ABG26186;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #26177.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS90373.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 20; SEQ ID NO 56545; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 194 AA;

Query Match 57.9%; Score 66; DB 22; Length 194;
Best Local Similarity 65.0%; Pred. No. 0.014;
Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db	134	NOKAVKETVGVSHTRHMDL	153
RESULT	8		
ABG04299			
XX	ABG04299	standard; Protein: 231 AA.	
XX	ABG04299;		
XX	13-FEB-2002	(first entry)	
XX	Novel human diagnostic protein #4290.		
DE	Human: chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder.		
XX	Homo sapiens.		
OS	WO200175067-A2.		
XX	11-OCT-2001.		
PD	30-MAR-2001; 2001WO-US08631.		
XX	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
XX	(HYSE-) HYSEQ INC.		
PA	Drmanac RT, Liu C, Tang YT;		
XX	WPI: 2001-639362/73.		
DR	N-PSDB: AAS68486.		
XX	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity		
XX			
PS	Claim 20; SEQ ID NO 34658; 103bp; English.		
XX			
CC	The invention relates to isolated polynucleotide (I) and		
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,		
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome		
CC	and gene mapping, and in recombinant production of (II). The		
CC	polynucleotides are also used in diagnostics as expressed sequence tags		
CC	for identifying expressed genes. (I) is useful in gene therapy techniques		
CC	to restore normal activity of (II) or to treat disease states involving		
CC	(II). (II) is useful for generating antibodies against it, detecting or		
CC	quantitating a polypeptide in tissue, as molecular weight markers and as		
CC	a food supplement. (II) and its binding partners are useful in medical		
CC	imaging of sites expressing (II). (I) and (II) are useful for treating		
CC	disorders involving aberrant protein expression or biological activity.		
CC	The polypeptide and polynucleotide sequences have applications in		
CC	diagnostics, forensics, gene mapping, identification of mutations		
CC	responsible for genetic disorders or other traits to assess biodiversity		
CC	and to produce other types of data and products dependent on DNA and		
CC	amino acid sequences. ABG00010-ABG30377 represent novel human		
CC	diagnostic amino acid sequences of the invention.		
CC	Note: The sequence data for this patent did not appear in the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX			
SO	Sequence	231 AA;	
Query Match	57.9%;	Score 66;	DB 22; Length 231;
Best Local Similarity	65.0%;	Pred. No. 0.017;	
Matches	13; Conservative	1; Mismatches	6; Indels 0; Gaps 0;
0Y	2 NSGDYKEDYGFATHTRADML 21		
	I I I I I I I I I I I I I		
db	35 NOKAVKETVGVSHTRHMDL 54		

RESULT 9	
ABG21816	
ID . ABG21816 standard; Protein: 241 AA.	
XX	
AC	ABG21816;
XX	
DT	18-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #21807.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
PA	(HYSE-) HYSEQ INC.
XX	
PI	Dmanac RT, Liu C, Tang YT;
XX	
DR	WPI: 2001-639362/73.
DR	N-PSDB; AAS86003.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
PS	Claim 20; SEQ ID No 52175; 103bp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ.	Sequence 241 AA;
Query Match	57.9%; Score 66; DB 22; Length 241;
Best Local Similarity	65.0%; Pred. No. 0.017;
Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0.	
OY	2 NSGDYKEDYGFAHITRADML 21
I	:
Db	35 NOKAYKETGYGVSHTRHMDL 54
RESULT 10	

XX	PN	WO200175067-A2.
XX	PD	11-OCT-2001.
XX	PF	30-MAR-2001; 2001WO-US08631.
XX	PR	31-MAR-2000; 2000US-0540217.
XX	PR	23-AUG-2000; 2000US-0649167.
XX	PA	(HYSE-) HKSEQ INC.
XX	PI	Dzmanac RT, Liu C, Tang YT;
XX	PI	WPI: 2001-639362/73.
XX	DR	N-PSDB: AAS74478.
XX	PT	New isolated polynucleotide and encoded polypeptides, useful in
XX	PT	diagnostics, forensics, gene mapping, identification of mutations
XX	PT	responsible for genetic disorders or other traits and to assess
XX	PT	biodiversity -
XX	PS	Claim 20; SEQ ID NO 40650; 103pp; English.
XX	CC	The invention relates to isolated polynucleotide (I) and
XX	CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX	CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX	CC	and gene mapping, and in recombinant production of (II). The
XX	CC	polynucleotides are also used in diagnostics as expressed sequence tags
XX	CC	for identifying expressed genes. (I) is useful in gene therapy techniques
XX	CC	to restore normal activity of (II) or to treat disease states involving
XX	CC	(II). (II) is useful for generating antibodies against it, detecting or
XX	CC	quantitating a polypeptide in tissue, as molecular weight markers and as
XX	CC	a food supplement. (II) and its binding partners are useful in medical
XX	CC	imaging of sites expressing (II). (I) and (II) are useful for treating
XX	CC	disorders involving aberrant protein expression or biological activity.
XX	CC	The polypeptide and polynucleotide sequences have applications in
XX	CC	diagnostics, forensics, gene mapping, identification of mutations
XX	CC	responsible for genetic disorders or other traits to assess biodiversity
XX	CC	and to produce other types of data and products dependent on DNA and
XX	CC	amino acid sequences. ABG00010-ABG30377 represent novel human
XX	CC	diagnostic amino acid sequences of the invention.
XX	CC	Note: The sequence data for this patent did not appear in the printed
XX	CC	specification, but was obtained in electronic format directly from WIPO
XX	CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	XX	Sequence 268 AA:
XX	SO	
XX	Query Match	57.9%; Score 66; DB 22; Length 268;
XX	Best Local Similarity	65.0%; Pred. No. 0.019;
XX	Matches 13; Conservative	1; Mismatches 6; Indels 0; Gaps 0;
XX	QY	2 NSCDYKEDYGFNHTRADML 21
XX	Db	1 III II : IIII III
XX		35 NQKAYKERYGVSHTRHDL 54
XX	RESULT 15	
XX	ABG04324	
XX	ID	ABG04324 standard; Protein: 276 AA.
XX	XX	ABG04324;
XX	AC	
XX	DT	13-FEB-2002 (first entry)
XX	DE	
XX	DE	Novel human diagnostic protein #4315.
XX	KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX	KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	OS	Homo sapiens.
XX	XX	
XX	XX	WO200175067-A2.

```

PD 11-OCT-2001.
PP
PE 30-MAR-2001; 2001W0-US08631.
PR -31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PP
PA (HYSEQ-) HYSEQ INC.
PI Dmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDB: AAS68511.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID NO 34683; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC and to produce other genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG03377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPRO
CC at ftp.wipro.int/pub/published_pct_sequences.
XX
XX Sequence 276 AA:
SQ
XX
XX Query Match 57.9%; Score 66; DB 22; Length 276;
XX Best Local Similarity 65.0%; Pred. No. 0.02;
XX Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0.
XX
XX 2 NSGDYKEDYGFAHITRADML 21
XX I III II :IIII III
XX 41 NOKAVKRTYGVSHTRHMDL 60

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:55:48 ; Search time 3.19672 Seconds
(without alignments)
500.202 Million cell updates/sec

Title: US-09-986-682B-2

Perfect score: 65

Sequence: 1 SVFDGSDGTQYQ 12

Scoring table:

BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

A.Geneseq_101002:*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	100.0	12	19	AAW48308
2	65	100.0	487	19	AAW48306
3	40	61.5	418	22	ABB62267
4	39	60.0	185	17	AAW03548
5	39	60.0	608	23	ABB97303
6	38.5	59.2	815	23	ABP35606
7	38	58.5	295	23	ABB95607
8	38	58.5	295	23	ABB85001
9	38	58.5	477	22	ABB58663
10	38	58.5	558	13	AAK26843

11	38	58.5	582	23	ABG61896
12	38	58.5	1192	22	ABB93388
13	37.5	57.7	204	22	ABG13290
14	37.5	57.7	1310	22	ABG10127
15	37	56.9	103	23	ABP34181
16	37	56.9	184	22	ABB69757
17	37	56.9	185	22	ABB59234
18	37	56.9	185	22	ABB69758
19	37	56.9	190	20	AAV13480
20	37	56.9	238	22	ABG24775
21	37	56.9	350	22	AAU60192
22	37	56.9	403	23	ABP30257
23	37	56.9	408	21	AAAB15898
24	37	56.9	416	23	ABP27335
25	37	56.9	587	22	AAU41035
26	37	56.9	597	22	ABB60160
27	37	56.9	629	23	ABB91126
28	37	56.9	769	23	AAU36282
29	37	56.9	811	22	AAU36282
30	37	56.9	899	22	ABB65488
31	37	56.9	899	22	ABB65489
32	37	56.9	960	22	ABB93642
33	37	56.9	1018	15	AAK63759
34	37	56.9	1018	17	AAK87028
35	37	56.9	1018	18	AAW06485
36	37	56.9	1286	23	ABB91476
37	37	56.9	2076	22	AAU34319
38	37	56.9	2186	22	AAU37320
39	36	55.4	16	22	AAAG9807
40	36	55.4	37	19	AAW75153
41	36	55.4	63	20	AAV11775
42	36	55.4	113	22	AAE07005
43	36	55.4	135	22	ABB58764
44	36	55.4	146	22	AAW64662
45	36	55.4	210	19	AAW39176

ALIGNMENTS

RESULT 1	
AAW48308	
ID	AAW48308 standard; Protein; 12 AA.
XX	
AC	AAW48308;
DT	04-JUN-1998 (first entry)
XX	
DE	Bacillus sp. V230 beta-fructofuranosidase internal fragment.
XX	
KW	Beta-fructofuranosidase; xylosyl fructoside; isomaltosyl fructoside;
KW	lactosucrose; fructosyltrehalose; antidiabetic; growth promoter;
KW	Bifidobacteria; mineral-absorption promoting activity; taste; texture;
KW	cosmetic; pharmaceutical; sweetener.
OS	Bacillus sp.
XX	
PN	EP812915-A2.
XX	
PD	17-DEC-1997.
XX	
PF	10-JUN-1997; 97EP-0304032.
XX	
PR	10-JUN-1996; 96JP-0170630.
XX	
PA	(HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
XX	
PI	Chaen H. Kubota M. Tsusaki K;
XX	
DR	WPI; 1998-034976/04.
XX	
PT	Bacillus beta-fructofuranosidase enzyme - useful for producing
PT	fructo-furanosylated saccharide(s) or alcohol(s)

XX Claim 1: Page 19; 32pp; English.
 PS This is an internal fragment of beta-fructofuranosidase from
 XX Bacillus sp. V230, an enzyme that has an optimum temperature of 50 deg. C
 CC at pH 6. This enzyme catalyses fructofuranosyl transfer reactions from a
 CC fructofuranosyl donor to a fructofuranosyl acceptor where the donor is
 CC sucrose, raffinose or erlose and the acceptor is selected from alcohols,
 CC sugar alcohols and saccharides having no beta-fructofuranosidic linkages,
 CC especially where the reaction product is xylosyl fructoside, erlose,
 CC isomaltosyl fructoside, lactosucrose or fructosyltrehalose.
 CC Such reaction products have a satisfactory taste and sweetness, a
 CC moderate viscosity and humectancy, an effective anticariogenic activity,
 CC growth promoting activity for Bifidobacteria, mineral-absorption
 CC promoting activity. The products can be used to improve the tastes and
 CC textures of food products, cosmetics and pharmaceuticals and are useful
 CC as sweeteners in the food and pharmaceutical industries.
 CC
 SQ Sequence 12 AA;
 Query Match 100.0%; Score 65; DB 19; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.00018;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SVFDGGDGTIVQ 12
 DB 1 SVFDGGDGTIVQ 12
 RESULT 2
 ID AAM48306 standard; Protein; 487 AA.
 AC AAM48306;
 DT 04-JUN-1998 (first entry)
 DE Bacillus sp. V230 beta-fructofuranosidase.
 KW Beta-fructofuranosidase; xylosyl fructoside; isomaltosyl fructoside;
 KW lactosucrose; fructosyltrehalose; anticariogenic; growth promoter;
 KW Bifidobacteria; mineral-absorption promoting activity; taste; texture;
 KW cosmetic; pharmaceutical; sweetener.
 OS Bacillus sp.
 FT Key Location/Qualifiers
 FT Peptide 1..32
 FT /label= signal_peptide
 FT Protein 33..487
 FT /label= beta-fructofuranosidase
 FT /note= "Claim 4"
 EP812915-A2.
 PD 17-DEC-1997.
 PF 10-JUN-1997; 97EP-0304032.
 PR 10-JUN-1996; 96JP-0170630.
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 PI Chaeen H, Kubota M, Tsusaki K;
 DR WPI: 1998-034976/04.
 DR N-PSDB; AAV17621.
 PT Bacillus beta-fructofuranosidase enzyme - useful for producing
 PT fructo:furanosylated saccharide(s) or alcohol(s)
 PS Claim 1; Pages 23-25; 32pp; English.

CC This is beta-fructofuranosidase from Bacillus sp. V230, an enzyme that
 CC has an optimum temperature of 50 deg. C at pH 6. It catalyses
 CC fructofuranosyl transfer reactions from a fructofuranosyl donor to a
 CC fructofuranosyl acceptor where the donor is sucrose, raffinose or erlose
 CC and the acceptor is selected from alcohols, sugar alcohols and
 CC saccharides having no beta-fructofuranosidic linkages,
 CC especially where the reaction product is xylosyl fructoside, erlose,
 CC isomaltosyl fructoside, lactosucrose or fructosyltrehalose.
 CC Such reaction products have a satisfactory taste and sweetness, a
 CC moderate viscosity and humectancy, an effective anticariogenic activity,
 CC growth promoting activity for Bifidobacteria, mineral-absorption
 CC promoting activity. The products can be used to improve the tastes and
 CC textures of food products, cosmetics and pharmaceuticals and are useful
 CC as sweeteners in the food and pharmaceutical industries.
 CC
 SQ Sequence 487 AA;
 Query Match 100.0%; Score 65; DB 19; Length 487;
 Best Local Similarity 100.0%; Pred. No. 0.01;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SVFDGGDGTIVQ 12
 DB 233 SVFDGGDGTIVQ 244
 RESULT 3
 ID ABB62267 standard; Protein; 418 AA.
 AC ABB62267;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 13593.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 FT WO200171042-A2.
 PD 27-SEP-2001.
 PF 23-MAR-2001; 2001WO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI: 2001-656860/75.
 DR N-PSDB; ABL06370.
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Disclosure; SEQ ID NO 13593; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (AB101840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 418 AA:

Query Match 61.5%; Score 40; DB 22; Length 413;

Best Local Similarity 70.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 VFDDGDDGVY 11

DB 13 LFDDGDMNVY 22

RESULT 4

ID AAM03548 standard; Protein: 185 AA.

XX AAM03548;

DT 21-APR-1997 (first entry)

DE ORF-2 protein sequence from BamHI fragment of HVT.

XX Herpes virus of turkey; open reading frame; ORF; homology; vector;
 KW avian herpes virus; recombinant viral vaccine; intergenic region; IBV;
 KW cytomegalovirus immediate early promoter; UL55 gene; repeat region; ILTV;
 KW antigen; infectious bursal disease virus; Marek's disease virus; MDV;
 KW infectious laryngotracheitis virus; avian anaemia virus; vaccination;
 KW infectious bronchitis virus; IBV; poultry; Gumboro disease;
 KW Newcastle disease.

XX Herpesvirus of turkey.

XX EP119864-A2.

PD 03-JUL-1996.

XX 28-DEC-1995; 95EP-0402970.

XX 30-DEC-1994; 94FR-0016017.

XX (INMR) RHONE MERIEUX SA.

XX Audonnet JCF, Bublout MJM, Dartell RJ, Duinat CV;

XX Laplace ELF, Riviere MAE;

XX WPI: 1996-364150/37.

XX N-PSDB: AAT39309.

XX Live recombinant avian vaccine - comprises herpes virus as vector

XX and having sequence encoding antigenic polypeptide inserted between

XX UL55 gene and repeat region

XX Example 4; Fig 1; 50pp; French.

XX This is the amino acid sequence of the protein encoded by the open
 CC reading frame 2 (ORF-2) from the BamHI I fragment from the herpes virus
 CC of turkeys (HVT). The protein has homology to the 185 amino acid protein
 CC encoded by the EHV-1 gene 3 ORF. The DNA fragment was isolated from HVT
 CC strain FC126 (J. Gen. Virol., 70:1789-1804 (1989)) and is used as a basis
 CC for generating live recombinant viral vaccines using the herpes virus DNA
 CC sequences as vectors. The fragment contains 3 intergenic regions into
 CC which genes encoding heterologous proteins to be expressed can be
 CC inserted, preferably under control of the cytomegalovirus immediate early
 CC (CMV-IE) promoter. The heterologous proteins can be used to vaccinate
 CC poultry against Gumboro disease (caused by IBV), Newcastle disease,
 CC Marek's disease, infectious bronchitis, infectious laryngotracheitis and
 CC avian anaemia.

XX Sequence 185 AA:

Query Match 60.0%; Score 39; DB 17; Length 185;

Best Local Similarity 60.0%; Pred. No. 78;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 VFDDGDDGVY 11

DB 74 IFEGDGSVY 83

RESULT 5

ID ABB97303 standard; Protein: 608 AA.

XX ABB97303;

XX 27-JUN-2002 (first entry)

XX Novel human protein SEQ ID NO: 571.

XX Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
 KW antiinfectivity; cerebroprotective; cytosolic; rheumatic; gene therapy;
 KW neuroprotective; antiparkinsonian; protein therapy; EST;
 KW expressed sequence tag.

XX Homo sapiens.

XX WO200222660-A2.

XX 21-MAR-2002.

XX 10-SEP-2001; 2001WO-US26015.

XX 11-SEP-2000; 2000US-0659671.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI: 2002-292408/33.

XX N-PSDB: ABN32489.

XX An isolated polynucleotide for treating diseases associated with its

XX encoded polypeptide such as cancer and multiple sclerosis -

XX Example 2; SEQ ID NO 571; 50pp; English.

XX The present invention provides the protein and coding sequences of 444
 CC novel human proteins. These were isolated from expressed sequences tags
 CC (ESTs). They can be used to stimulate cell growth, to regulate
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
 CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
 CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
 CC Parkinson's disease. The present sequence is a protein of the invention.

XX Sequence 608 AA:

Query Match 60.0%; Score 39; DB 23; Length 608;

Best Local Similarity 54.5%; Pred. No. 2.9e+02;

Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 VFDDGDDGVYQ 12

DB 310 LFEGGKGSVFQ 320

RESULT 6

ID ABB973606 standard; Protein: 815 AA.

XX ABB973606;

XX ABB973606;

DT		24-JUL-2002	(first entry)
DE		Fungal ZBC protein sequence #32.	
XX			
XX		Secondary metabolite; fungus; ZBC gene; zinc binuclear cluster protein;	
XX		antibacterial; beta-lactam; anti-hypercholesterolaemic; lovastatin;	
KW		mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;	
KM		angio genesis inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;	
KV		fungal toxin; cell surface receptor; plant growth regulator; pigment;	
KX		insecticide; antineoplastic.	
OS		Unidentified.	
PN		WO200224865-A2.	
PD		28-MAR-2002.	
PF		19-SEP-2001; 2001WO-US29288.	
PR		19-SEP-2000; 2000US-233564P.	
PA		(MICR-) MICROBIA INC.	
PI	Holtzman D, Madden K, Maxon M, Sherman A:		
DR	WPI: 2002-352005/38.		
DR	N-PsDB: ABN79795.		
PT	New method for improving the production of a secondary metabolite e.g.		
PT	antineoplastic agent, ergot alkaloid from a fungus involves modulation		
PT	of the expression of at least one zinc binuclear cluster protein gene		
PS	Disclosure: SEQ ID 76; 49pp + sequence listing; English.		
CC	The invention relates to improving the production of a secondary		
CC	metabolite by a fungus. This involves modulating the expression of at		
CC	least one ZBC (zinc binuclear cluster protein) gene in a manner to		
CC	improve the yield of the secondary metabolite. Methods of the invention		
CC	may be used for improving the production of the secondary metabolite e.g.		
CC	antibacterial (such as beta-lactam), an anti-hypercholesterolaemic (such		
CC	as lovastatin or mevastatin), an immunosuppressant (such as cyclosporin A),		
CC	an ergot alkaloid (such as ergotamine), an angiogenesis inhibitor (such		
CC	as ovalicin), a glucan synthase inhibitor, gliotoxin family of compounds,		
CC	a fungal toxin, a modulator of cell surface receptor signalling, a plant		
CC	growth regulator, a pigment, an insecticide, or an antineoplastic		
CC	compound. The method results in a decrease in fermentor run-time, a		
CC	decrease in the size of the fermentor required for the production of		
CC	the biomass amounts of the secondary metabolite, or a decrease in the		
CC	waste that must be handled in downstream processing. The sequences given		
CC	in records ABP35575-ABP35722 represent ZBC proteins.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained directly from WIPO at		
CC	ftp.wipo.int/pub/published_pcl_sequences.		
XX			
Sequence	815 AA:		
Query Match	59.2%;	Score 38.5; DB 23; Length 815;	
Best Local Similarity	69.2%;	Pred. NO. 4.8e+02;	
Matches 9; Conservative	1;	Mismatches 2; Indels 1; Gaps 1;	
OY	1 SVFDGGDGT-V-YQ 12		
	:		
Db	680 SVFDGNITGTYQ 692		
RESULT 7			
ID	ABB95607		
AC	ABB95607 standard; Protein: 295 AA.		
XX	ABB95607;		
XX			

DT	19-JUL-2002	(first entry)	
XX			
DE	Human angiogenesis related protein PRO28631	SEQ ID NO: 370.	
XX			
KW	Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;		
KW	atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder		
KW	cardiac; cytosolic; antiangiogenic; hypotensive; vulnerary;		
KW	antiarteriosclerotic.		
XX			
OS	Homo sapiens.		
XX			
PN	MO200208284-A2.		
XX			
PD	31-JAN-2002.		
XX			
PF	09-JUL-2001; 2001WO-US21735.		
XX			
XX	20-JUL-2000; 2000US-219556P.		
PR	25-JUL-2000; 2000US-220624P.		
PR	25-JUL-2000; 2000US-220664P.		
PR	28-JUL-2000; 2000WO-US20710.		
PR	02-AUG-2000; 2000US-222695P.		
PR	17-AUG-2000; 2000US-0643657.		
PR	23-AUG-2000; 2000WO-US23522.		
PR	24-AUG-2000; 2000WO-US23328.		
PR	07-SEP-2000; 2000US-230978P.		
PR	15-SEP-2000; 2000US-000000P.		
PR	18-SEP-2000; 2000US-0664610.		
PR	18-SEP-2000; 2000US-0663550.		
PR	24-OCT-2000; 2000US-242922P.		
PR	08-NOV-2000; 2000US-070923P.		
PR	08-NOV-2000; 2000WO-US30552.		
PR	10-NOV-2000; 2000WO-US30873.		
PR	01-DEC-2000; 2000WO-US32678.		
PR	20-DEC-2000; 2000US-0747259.		
PR	20-DEC-2000; 2000WO-US34956.		
PR	22-JAN-2001; 2001US-0767609.		
PR	28-FEB-2001; 2001US-0796498.		
PR	28-FEB-2001; 2001WO-US06520.		
PR	01-MAR-2001; 2001WO-US06666.		
PR	09-MAR-2001; 2001US-0802706.		
PR	14-MAR-2001; 2001US-0808689.		
PR	22-MAR-2001; 2001US-0816744.		
PR	05-APR-2001; 2001US-0828366.		
PR	10-MAY-2001; 2001US-0854208.		
PR	10-MAY-2001; 2001US-0854280.		
PR	25-MAY-2001; 2001US-0866028.		
PR	25-MAY-2001; 2001US-0866034.		
PR	25-MAY-2001; 2001WO-US17092.		
PR	30-MAY-2001; 2001US-0870574.		
PR	30-MAY-2001; 2001WO-US17443.		
PR	01-JUN-2001; 2001WO-US17800.		
PR	20-JUN-2001; 2001WO-US19692.		
PR	28-JUN-2001; 2001WO-US00000.		
XX			
PA	(GETH) GENENTECH INC.		
PA	(BAKE/) BAKER K P.		
PA	(FERR/) FERRARA N.		
PA	(GERB/) GERBER H.		
PA	(GERR/) GERRITSEN M E.		
PA	(GODD/) GODDARD A.		
PA	(GODO/) GODOWSKI P J.		
PA	(GURN/) GURNEY A L.		
PA	(HILL/) HILLAN K J.		
PA	(MARS/) MARSTERS S A.		
PA	(PANT/) PAN J.		
PA	(PRON/) PRONI N F.		
PA	(STEP/) STEPHAN J F.		
PA	(MATA/) WATANABE C K.		
PA	(WILL/) WILLIAMS P M.		
PA	(WOOD/) WOOD W I.		
XX			
XX	Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;		

PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
 PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
 XX MPI: 2002-171999/22.
 DR N-PSDB: ABL95745.

XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,
 PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
 PT infarction), endothelial or angiogenic disorders in a mammal -
 XX
 PS Claim 11; Fig 370; 567pp; English.

XX The present invention provides the protein and coding sequences of human
 CC PRO proteins. These are useful for treating or diagnosing a
 CC cardiovascular, endothelial or angiogenic disorder, including cardiac
 CC hypertrophy, trauma, cancer, age-related macular degeneration,
 CC atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
 CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
 CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound
 CC healing. The present sequence is a PRO protein of the invention.

XX Sequence 295 AA;
 SQ

Query Match 58.5%; Score 38; DB 23; Length 295;
 Best Local Similarity 75.0%; Pred. No. 1.9e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 FDGGDGTV 10
 |||||
 Db 58 FDGGDATI 65

RESULT 8
 ABB85001
 ID ABB85001 standard; Protein: 295 AA.

XX ABB85001;
 AC
 XX
 DT 16-MAY-2002 (first entry)
 XX
 DE Human PRO28631 protein sequence SEQ ID NO:370.
 XX
 DE Human; angiogenesis; cardiant; cytosolic; antilangiogenic; hypotensive;
 KW vulnerability; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
 KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;
 KW angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
 KW age-related macular degeneration; arterial restenosis; angina;
 KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;
 KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
 KW wound healing; chromosome mapping; gene mapping.

XX Homo sapiens.
 OS
 XX WO200200690-A2.
 PN
 XX
 PD 03-JAN-2002.
 XX
 PF 20-JUN-2001; 2001WO-US19692.
 XX
 XX 23-JUN-2000; 2000US-213637P.
 PR 20-JUL-2000; 2000US-219556P.
 PR 25-JUL-2000; 2000US-220624P.
 PR 28-JUL-2000; 2000US-220664P.
 PR 02-AUG-2000; 2000US-222695P.
 PR 17-AUG-2000; 2000US-0643657.
 PR 23-AUG-2000; 2000WO-US23522.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 07-SEP-2000; 2000US-230978P.
 PR 18-SEP-2000; 2000US-0664610.
 PR 24-OCT-2000; 2000US-0665350.
 PR 08-NOV-2000; 2000US-242922P.
 XX 08-NOV-2000; 2000US-0709238.

PR 08-NOV-2000; 2000WO-US30952.
 PR 10-NOV-2000; 2000WO-US30873.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000US-0747259.
 PR 20-DEC-2000; 2000US-0747259.
 PR 22-JAN-2001; 2000US-0767609.
 PR 28-FEB-2001; 2001US-0796498.
 PR 28-FEB-2001; 2001WO-US06520.
 PR 01-MAR-2001; 2001WO-US06665.
 PR 09-MAR-2001; 2001US-0802706.
 PR 14-MAR-2001; 2001US-0808689.
 PR 22-MAR-2001; 2001US-0816744.
 PR 05-APR-2001; 2001US-0828366.
 PR 10-MAY-2001; 2001US-0854208.
 PR 10-MAY-2001; 2001US-0854280.
 PR 25-MAY-2001; 2001US-0866028.
 PR 25-MAY-2001; 2001US-0866034.
 PR 25-MAY-2001; 2001WO-US17092.
 PR 30-MAY-2001; 2001US-0870574.
 PR 30-MAY-2001; 2001WO-US17443.
 PR 01-JUN-2001; 2001WO-US17800.
 XX
 XX (GETH) GENENTECH INC.
 PA
 XX Baker KP, Ferrara N, Gerber H, Gertlisen ME, Goddard A;
 PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
 PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
 XX MPI: 2002-090516/12.
 DR N-PSDB: ABL88256.

XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,
 PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
 PT infarction), endothelial or angiogenic disorders in a mammal -
 XX
 PS Claim 11; Fig 370; 565pp; English.

XX ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to
 CC ABB85003. The PRO proteins and polynucleotides have cardiant, cytosolic,
 CC antilangiogenic, hypotensive, vulnerability and antiarteriosclerotic
 CC activities, and can be used in gene therapy. The PRO polynucleotides,
 CC proteins, agonists and antagonists are useful for treating or diagnosing
 CC a cardiovascular, endothelial or angiogenic disorder in a mammal,
 CC e.g. cardiac hypertrophy, trauma, cancer, age-related macular
 CC degeneration, atherosclerosis, hypertension, arterial restenosis,
 CC rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis,
 CC lymphangitis, tumour angiogenesis (such as breast carcinoma and liver
 CC carcinoma) and wound healing. The PRO polynucleotides have applications
 CC in molecular biology, including use as hybridisation probes, and in
 CC chromosome and gene mapping. ABL88259 to ABL88267 represent primers and
 CC probes used in the exemplification of the present invention.

XX Sequence 295 AA;
 SQ

Query Match 58.5%; Score 38; DB 23; Length 295;
 Best Local Similarity 75.0%; Pred. No. 1.9e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 FDGGDGTV 10
 |||||
 Db 58 FDGGDATI 65

RESULT 9
 ABB58863
 ID ABB58863 standard; Protein: 477 AA.
 XX
 AC ABB58863;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 3381.
 XX

KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
XX	
XX	Drosophila melanogaster.
OS	
XX	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
XX	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
XX	WPI: 2001-656860/75.
DR	N-PSDB; ABL02966.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
PS	Disclosure: SEQ ID NO 3381; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC	sequences (AB57737-AB572072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SO	Sequence 477 AA;
QY	1 SVFDGSDGTGY 11
Db	189 SMFLGFDCTYV 199
XX	
XX	RESULT 10
XX	AAAR26843
XX	AAAR26843 standard; Protein: 558 AA.
XX	
XX	AAAR26843;
XX	
DT	11-FEB-1993 (first entry)
XX	
DE	GDP dissociation stimulatory protein.
XX	
KW	Guanosine 5'-diphosphate; mass production.
XX	
OS	Not known.
XX	
PN	JP04211700-A.
PD	
XX	03-AUG-1992.
XX	
PF	08-FEB-1991; 91JP-0018105.
XX	
PR	24-MAY-1990; 90JP-0134480.
XX	
PA	(MITU) MITSUBISHI KASEI CORP.

XX	WPI: 1992-305013/37.
DR	N-PSDB: AAO28232.
XX	
PT	High mol. wt. mass producible protein - accelerates dissociation
PT	of guanosine 5'-di:phosphate from complex of GDP and GDP protein
PT	binding type low mol guanosine 5'-tri:phosphate
XX	
PS	Claim 1; Fig 1; 17pp: Japanese.
XX	
CC	The sequence is that of a protein which accelerates dissociation of
CC	guanosine 5'-di:phosphate (GDP) from a complex of GDP and GDP protein
CC	binding type low molecular guanosine 5'-triphosphate (GTP) binding
CC	protein. It may be mass produced using genetic engineering techniques.
XX	
SQ	Sequence 558 AA:
Query Match	58.5%; Score 38; DB 13; Length 558;
Best Local Similarity	54.5%; Pred. No. 3.9e+02;
Matches 6; Conservative	3; Mismatches 2; Indels 0; Gaps 0;
QY	2 VFDDGGDTGVYQ 12
	: :
DB	260 LFEQGGKGNVFQ 270
RESULT 11	
ABG61896	
ID	ABG61896 standard; Protein; 582 AA.
XX	
AC	ABG61896;
XX	
DT	15-AUG-2002 (first entry)
XX	
DE	Prostate cancer-associated protein #97.
XX	
KW	Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.
XX	
OS	Mammalia.
XX	
PN	WO200230268-A2.
XX	
PD	18-APR-2002.
XX	
PF	12-OCT-2001; 2001WO-US32045.
XX	
PR	13-OCT-2000; 2000US-0687576.
PR	08-DEC-2000; 2000US-0733288.
PR	08-DEC-2000; 2000US-0733742.
PR	24-JAN-2001; 2001US-263957P.
PR	16-MAR-2001; 2001US-276791P.
PR	16-MAR-2001; 2001US-276888P.
PR	06-APR-2001; 2001US-281922P.
PR	24-APR-2001; 2001US-286214P.
PR	30-APR-2001; 2001US-0847046.
PR	04-MAY-2001; 2001US-288589P.
XX	
PA	(EOSB-) EOS BIOTECHNOLOGY INC.
XX	
PI	Gish KC, Mack DH, Wilson KE, Afar D, Heyezi P;
XX	
DR	WPI: 2002-471335/50.
DR	N-PSDB: ABK92212.
XX	
PT	Detecting a prostate cancer-associated transcript in a cell in a
PT	patient, useful for diagnosing prostate cancer (PC) or screening
PT	modulators of PC, by determining if prostate cancer-associated genes
PT	are expressed in a prostate tissue
XX	
PS	Claim 27; Page 383; 436pp: English.
XX	
CC	The present invention relates to methods of detecting a prostate
CC	cancer-associated transcript in a cell from a patient. The method

Matches 8; Conservative 1; Mismatches 2; Indels 1; Gaps 1;
QY 1 SVEPDGDTFYQ 12
|:|:| | | | |
Db 176 SIFD-GDCKTYQ 186

RESULT 14
ABG10127
ID ABG10127 standard; Protein: 1310 AA.
XX

AC ABG10127;
XX

DT 13-FEB-2002 (first entry)
XX

DE Novel human diagnostic protein #10118.
XX

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX

OS Homo sapiens.
XX

PN W0200175067-A2.
XX

PD 11-OCT-2001.
XX

PF 30-MAR-2001; 2001WO-US08631.
XX

PR 31-MAR-2000; 2000US-0540217.
XX

PR 23-AUG-2000; 2000US-0649167.
XX

PA (HYSE-) HYSEQ INC.
XX

PI Drmanac RT, Liu C, Tang YT;
XX

DR WPI: 2001-639362/73.
XX

DR N-PSDB: AAS74314.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID NO 40486; 103pp; English.
XX

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 1310 AA;
XX

Query Match 57.7%; Score 37.5; DB 22; Length 1310;
Best Local Similarity 66.7%; Pred. No. 1.2e+03;
Matches 8; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 SVEPDGDTFYQ 12
|:|:| | | | |
Db 1282 SIFD-GDCKTYQ 1292

RESULT 15
ABP34181
ID ABP34181 standard; Protein: 103 AA.
XX

AC ABP34181;
XX

DT 08-JUL-2002 (first entry)
XX

DE Human glycoprotein-like ORF3154 protein, SEQ ID NO:6308.
XX

KW Human: ORF; open reading frame; ORF; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibin; chemotactic; chemokine; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW candid; hypotensive; antihypertoid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide.
XX

OS Homo sapiens.
XX

PN W0200190366-A2.
XX

PD 29-NOV-2001.
XX

PF 24-MAY-2001; 2001WO-US17076.
XX

PR 24-MAY-2000; 2000US-206690P.
XX

PA (CURA-) CURAGEN CORP.
XX

PI Leach MD, Shinkels RA;
XX

DR WPI: 2002-106200/14.
XX

DR N-PSDB: ABN78207.
XX

PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation -

PS Claim 10; Page 1826-1827; 2508pp; English.
XX

CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC ABN79587 represent CDNA's encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORP1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antiinfective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,

CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases.
XX

SQ Sequence 103 AA;

Query Match 56.9%; Score 37; DB 23; Length 103;
Best Local Similarity 77.8%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 DGGDGTWYQ 12
|||||
Db 66 DGGDGVVSQ 74

Search completed: June 16, 2003, 12:06:46
Job time : 4.19672 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:01:28 ; Search time 3.02459 Seconds
(without alignments)
817.488 Million cell updates/sec

Title: US-09-986-682B-2
Perfect score: 65
Sequence: 1 SVFDGSDGTIVYQ 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	100.0	487	2	082854
2	59	90.8	489	16	097181
3	54	83.1	499	2	0925E5
4	45	69.2	800	2	09F8X1
5	44	67.7	125	16	08RI28
6	44	67.7	888	17	0977D4
7	43	66.2	330	16	09A934
8	42	64.6	926	5	020176
9	41	63.1	239	10	09F1C9
10	41	63.1	263	5	09B1L7
11	41	63.1	295	16	09KF21
12	41	63.1	380	16	09ZBK8
13	41	63.1	2658	13	090WF0
14	40	61.5	225	16	086806
15	40	61.5	418	5	09VTD9
16	40	61.5	469	16	050499

17	39	60.0	108	10	09M904	09m9u4 arabidopsis
18	39	60.0	175	10	08RXS1	08rxs1 arabidopsis
19	39	60.0	185	12	09DGT5	09dgt5 melegrid h
20	39	60.0	212	10	09FTK1	09ftk1 oryza sativ
21	39	60.0	346	16	09RD67	09rd67 streptomyce
22	39	60.0	406	4	09B0W9	09b0w9 homo sapien
23	39	60.0	444	10	09XHF2	09xhf2 zea mays (m
24	39	60.0	523	8	09ZY25	09zy25 pedionomas
25	39	60.0	558	11	092105	092105 mus musculu
26	39	60.0	607	4	09N2A8	09n2a8 homo sapien
27	39	60.0	607	4	09NFM2	09nfm2 homo sapien
28	39	60.0	607	13	09B0X6	09b0x6 homo sapien
29	39	60.0	607	13	093614	093614 xenopus lae
30	39	60.0	1299	16	08XMT7	08xmt7 raltsonia s
31	39	60.0	1420	10	09XEJ3	09xej3 oryza sativ
32	39	60.0	2185	9	08W6J4	08w6j4 sinorhizobi
33	39	60.0	2454	5	08T2G3	08t2g3 dictyosteli
34	39	60.0	4283	11	09ERV0	09erv0 ratius norv
35	39	60.0	4293	11	008852	008852 mus musculu
36	38.5	59.2	647	17	09HOM8	09hgm8 halobacteri
37	38.5	59.2	815	3	059744	059744 schizosacch
38	38	58.5	146	10	023043	023043 arabidopsis
39	38	58.5	214	7	09TNV6	09tnv6 oncorhynch
40	38	58.5	295	11	08QZV2	08qzv2 mus musculu
41	38	58.5	358	2	09X5K8	09x5k8 streptomyce
42	38	58.5	364	2	007862	007862 streptomyce
43	38	58.5	364	16	09RJY2	09rjy2 streptomyce
44	38	58.5	398	16	067179	067179 aquifex aeo
45	38	58.5	418	16	09WYC7	09wyc7 thermotoga

ALIGNMENTS

RESULT 1

ID 082854 PRELIMINARY: PRT: 487 AA.
AC 082854;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Beta-fructofuranosidase.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V230;
RA Tsusaki K.;
RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-V230;
RA Kurimoto M.;
RT "Cloning and sequencing of beta-fructofuranosidase gene from Bacillus
RT sp. V230."
RL Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB010272; BAA32083.1; -;
DR InterPro; IPR003469; Glyco_hydro.68.
DR Pfam; PF02435; Glyco_hydro.68; 1.
SO SEQUENCE 487 AA; 53412 MW; 7F52A47921824AE3 CRC64;

Query Match 100.0%; Score 65; DB 2; Length 487;
Best local Similarity 100.0%; Pred. No. 0.017;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SVFDGSDGTIVYQ 12
|||||
Db 233 SVFDGSDGTIVYQ 244

RESULT 2

097181 PRELIMINARY; PRT: 489 AA.
 ID 097181;
 AC 097181;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 GN CAC1772.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 NCBI_TaxID=1488;
 RX MEDLINE=11466286;
 RA Gibson R., Lee H.M., Dupois J., Qiu D., Hittl J., Wolf Y.I.,
 Tatusov R.L., Sabath F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 bacterium Clostridium acetobutylicum.";
 RL J. Bacteriol. 183:4823-4836(2001).
 DR EMBL: AE007686; AAK79737.1;
 DR InterPro: IPR003469; Glyco_hydro.68.
 DR Pfam: PF02435; Glyco_hydro.68; 1;
 KW Complete proteome.
 SQ SEQUENCE 489 AA; 54553 MW; 9002CB364F1D3CD8 CRC64;

Query Match 90.8%; Score 59; DB 16; Length 489;
 Best Local Similarity 83.3%; Pred. No. 0.16;
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SVFDGDDGTVO 12
 |||||:||||
 DB 229 SVFDGDDGTVO 240

RESULT 3
 0925E5 PRELIMINARY; PRT: 499 AA.
 ID 0925E5;
 AC 0925E5;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Levensucrase.
 GN SABC.
 OS Paenibacillus polymyxa (Bacillus polymyxa).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Paenibacillaceae; Paenibacillus.
 NCBI_TaxID=1406;
 RX MEDLINE=11200435;
 RA Bezzate S., Aymeric S., Chamber R., Czarnes S., Berge O., Heulin T.;
 RT "Disruption of the Paenibacillus polymyxa levensucrase gene impairs
 its ability to aggregate soil in the wheat rhizosphere.";
 RL Environ. Microbiol. 2:333-342(2000).
 DR EMBL: AJ133737; CAB39327.1;
 DR InterPro: IPR003469; Glyco_hydro.68.
 DR Pfam: PF02435; Glyco_hydro.68; 1;
 SQ SEQUENCE 499 AA; 55317 MW; E646CD986292336C CRC64;

Query Match 83.1%; Score 54; DB 2; Length 499;
 Best Local Similarity 66.7%; Pred. No. 1.1;
 Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SVFDGDDGTVO 12
 |||||:||||
 DB 231 SVFDGDDGTVO 242

RESULT 4
 09F8X1 PRELIMINARY; PRT: 800 AA.
 ID 09F8X1;
 AC 09F8X1;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE N,N'-diacetylchitobiose phosphorylase.
 GN CHBP.
 OS Vibrio furnissii.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 NCBI_TaxID=29494;
 RX MEDLINE=10913116;
 RA Park J.K., Keyhani N.O., Roseman S.;
 RT "Chitin catabolism in the marine bacterium Vibrio furnissii.
 IDENTIFICATION, MOLECULAR CLONING, AND CHARACTERIZATION OF A N,N'-
 DIACETYLCHITOBIOSE PHOSPHORYLASE.";
 RL J. Biol. Chem. 275:33077-33083(2000).
 DR EMBL: AF230379; AAG23740.1;
 SQ SEQUENCE 800 AA; 89666 MW; 6775DB23C051122A CRC64;

Query Match 69.2%; Score 45; DB 2; Length 800;
 Best Local Similarity 88.9%; Pred. No. 5.1;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 DGGDGTVO 12
 |||||
 DB 456 DGGDGTVO 464

RESULT 5
 08R128 PRELIMINARY; PRT: 125 AA.
 ID 08R128;
 AC 08R128;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein FNI808.
 GN FNI808.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteria; Fusobacterium.
 NCBI_TaxID=76856;
 RX MEDLINE=11889109;
 RA Kapural V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Gotsman E., Bernal A.,
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Heselkorn R.,
 RA Fongstein M., Kyridides N., Overbeek R.;
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium
 nucleatum strain ATCC 25586.";
 RL J. Bacteriol. 184:2005-2018(2002).
 DR EMBL: AE010483; AAL3907.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 125 AA; 13549 MW; 661F5D709B72ABC5 CRC64;

Query Match 67.7%; Score 44; DB 16; Length 125;
 Best Local Similarity 72.7%; Pred. No. 10;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SVFDGDDGTVO 11
 |||||
 DB 11 SIDDGDDGTVO 21

RESULT 6
 0977D4 PRELIMINARY; PRT: 888 AA.
 ID 0977D4;
 AC 0977D4;

DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Hypothetical protein ST0012.
 GN ST0012.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=111955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JCM 10545 / 7;
 RX PubMed-11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 Sakane M., Baba S.-I., Anai A., Kosugi H., Hosoyama A., Fukui S.,
 Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
 Aoki K.-I., Masuda S., Yanagii M., Nishimura J., Yamagishi A.,
 Oshima T., Kikuchi H.;
 RA "Complete genome sequence of an aerobic thermocidophilic
 RT Crenarchaeon, Sulfolobus tokodaii strain7.";
 RL DNA Res. 8:123-140(2001).
 DR EMBL, AP000981; BAB64960.1; -
 DR InterPro: IPR002819; HD.
 DR InterPro: IPR003607; ME_Plpase_HDC.
 DR Pfam: PF01966; HD; 1.
 DR SMART: SM00471; HDC; 1.
 DR TIGRFAMS: TIGR00277; HDG; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 888 AA; 101756 MW; 008CC23F450C1AF CRC64;

Query Match 67.7%; Score 44; DB 17; Length 888;
 Best Local Similarity 72.7%; Pred. NO. 83;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 VFDDGDTVYQ 12
 1111111111
 DB 15 VFDDGDTVYQ 25

RESULT 7
 Q9A934 PRELIMINARY; PRT; 330 AA.
 AC Q9A934;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Hypothetical protein CC1160.
 GN CC1160.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 19089 / CB15;
 RX MEDLINE-21173698; PubMed-11259647;
 RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 Eichen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 Polocka I., Nelson W.C., Newton A.S., Stephens C., Phadke N.D., Ely B.,
 DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty K.,
 Uterback T., Tran K., Wolf A., Vamathavan J., Ermoiaeva M., White O.,
 Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RA "Complete genome sequence of Caulobacter crescentus.";
 RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 RL EMBL, AE005794; AAK3144.1; -
 DR TIGR: CC1160; -
 DR InterPro: IPR001206; DAGKC.
 DR Pfam: PF00781; DAGKC; 1.
 DR SMART: SM00046; DAGKC; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 330 AA; 35352 MW; 4112B72A9615E7C9 CRC64;

Query Match 66.2%; Score 43; DB 16; Length 330;
 Best Local Similarity 88.9%; Pred. NO. 42;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VFDDGDTV 10
 1111111111
 DB 80 VFDDGDTV 88

RESULT 8
 Q20176 PRELIMINARY; PRT; 926 AA.
 ID Q20176;
 AC Q20176;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Hypothetical 105.3 kDa protein.
 GN F38E9.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE-99069613; PubMed-9851916;
 RA None.
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Gattung S., Wu X.;
 RT "The sequence of C. elegans cosmid F38E9.";
 RT Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RT "Direct Submission.";
 RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 9 CUB DOMAINS.
 DR EMBL: U46668; AAA93348.2; -
 DR HSP: P00736; IAPQ.
 DR InterPro: IPR001506; Astacin.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01400; Astacin; 2.
 DR Pfam: PF00431; CUB; 8.
 DR Pfam: PF00008; EGF; 4.
 DR SMART: SM00042; CUB; 9.
 DR SMART: SM00181; EGF; 2.
 DR SMART: SM00179; EGF_Ca; 4.
 DR SMART: SM00235; Zmnc; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 4.
 DR PROSITE: PS01180; CUB; 8.
 DR PROSITE: PS01186; EGF_2; 4.
 DR PROSITE: PS01187; EGF_Ca; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN.1.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation;
 KW Hypothetical protein; Repeat.
 SQ SEQUENCE 926 AA; 105287 MW; A76F04B28B77D00 CRC64;

Query Match 64.6%; Score 42; DB 5; Length 926;
 Best Local Similarity 77.8%; Pred. NO. 18+02;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

DB      751 SVYDGGDGS 759

RESULT 9
Q9FLC9 PRELIMINARY: PRT: 239 AA.
ID Q9FLC9
AC Q9FLC9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Genomic DNA, chromosome 5, TAC clone:K18123 (At5g05250/K18123_5).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98344145; PubMed=9679202;
RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
RT features of the regions of 1,381,565 bp covered by twenty one
RT physically assigned P1 and TAC clones."
RL DNA Res. 5:131-145(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones."
RT Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
RL EMBL; AB010692; BAB09968.1;
DR EMBL; AY054291; AAL06949.1;
SQ SEQUENCE 239 AA: 25863 MW: 483C313F537EE2CA CRC64;

Query Match 63.1%; Score 41; DB 10; Length 239;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PDGGDGT 9
DB 166 PDGGDGT 172

RESULT 10
Q9BL17 PRELIMINARY: PRT: 263 AA.
ID Q9BL17
AC Q9BL17;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Fibroin L-chain.
GN FIB-L.
OS Dendrolimus spectabilis (pine moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Bombycoidea; Lasiocampidae; Dendrolimus.
OX NCBI_TaxID=155323;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21167527; PubMed=11267905;
RA Tanaka K., Mizuno S.;
RT "Homologues of fibroin L-chain and P25 of Bombyx mori are present in
RT Dendrolimus spectabilis and Papilio xuthus but not detectable in
RT Antherea yamamai."

```

```

RL Insect Biochem. Mol. Biol. 31:665-677(2001).
DR EMBL; AB001822; BAB39501.1;
SQ SEQUENCE 263 AA: 27501 MW: FFA2BF0943F4930C CRC64;

Query Match 63.1%; Score 41; DB 5; Length 263;
Best Local Similarity 60.0%; Pred. No. 70;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 VFDDGDTVY 11
DB 52 LFDGDDNMT 61

RESULT 11
Q9KF21 PRELIMINARY: PRT: 295 AA.
ID Q9KF21
AC Q9KF21;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein BH0676.
GN BH0676.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kunara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001509; BAB04393.1;
DR InterPro; IPR005218; Cons_Hypoth147.
DR InterPro; IPR001206; DAGC.
DR InterPro; IPR003622; DAG_kin_cat.
DR Pfam; PF00781; DAGKc_1.
DR ProDom; PD005043; DAG_kin_cat; 1.
DR SMART; SM00046; DAGKc_1.
DR TIGRFAMs; TIGR00147; TIGR00147; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 295 AA: 32456 MW: CFD93DDE5F40B290 CRC64;

Query Match 63.1%; Score 41; DB 16; Length 295;
Best Local Similarity 75.0%; Pred. No. 79;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 GDDGTVYQ 12
DB 65 GDDGTVY 72

RESULT 12
Q9ZBK8 PRELIMINARY: PRT: 380 AA.
ID Q9ZBK8
AC Q9ZBK8;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein SC06466.
GN SC06466 OR SC9C7.02.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;

```

RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RN Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapalate D., Elchner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL; AL035161; CAA22714.1;
 DR InterPro: IPR004381; Cons_hypoth45.
 DR InterPro: IPR003747; Glycerate_kinase.
 DR Pfam: PF02595; DUF168; 1.
 DR TIGRFAMs: TIGR00045; Cons_hypoth45; 1.
 DR Hypothetical protein.
 KW SEQUENCE 380 AA; 37816 MW; 7131A7C8735124EB CRC64;
 SQ
 Query Match 63.1%; Score 41; DB 16; Length 380;
 Best Local Similarity 88.9%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 VFDDGDTV 10
 Db 45 VADGGDTV 53
 RESULT 13
 Q90WFO PRELIMINARY; PRT; 2658 AA.
 ID Q90WFO;
 AC Q90WFO;
 DT 01-DEC-2001 (TRENBLREL. 19, Created)
 DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
 DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
 DE CGABP260.
 OS Gallus gallus (Chicken).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 NCBI_TaxID=9031;
 RX NCBI [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SKELETAL MUSCLE;
 RA Oshima K., Morikawa N., Hashimoto Y., Kuribayashi T., Tachikawa M.,
 RA Terasaki A.G., Ohashi K.;
 RT "Full cDNA sequence of cgABP260.";
 RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AB056475; BAB63944.1;
 DR InterPro: IPR001589; Actinbind_actin.
 DR InterPro: IPR001715; Calponin-like.
 DR InterPro: IPR001298; Filamin.
 DR Pfam: PF00307; CH; 2.
 DR Pfam: PF00630; Filamin; 24.

DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
 DR PROSITE: PS50021; CH; 2
 DR PROSITE: PS50194; FILAMIN_REPEAT; 20
 SQ SEQUENCE 2658 AA; 280492 MW; 899436B3257F778 CRC64;
 Query Match 63.1%; Score 41; DB 13; Length 2658;
 Best Local Similarity 88.9%; Pred. No. 8.1e+02;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SVFDGDCGT 9
 Db 1292 SVSDGDCGT 1300
 RESULT 14
 Q96806 PRELIMINARY; PRT; 225 AA.
 ID Q96806;
 AC Q96806;
 DT 01-NOV-1998 (TRENBLREL. 08, Created)
 DT 01-NOV-1998 (TRENBLREL. 08, Last sequence update)
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
 DE Hypothetical protein SC05746.
 DE SC05746 OR SC7C7.01.
 OS Streptomyces coelicolor.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Streptomycetaceae; Streptomyces.
 NCBI_TaxID=1902;
 RX NCBI [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Harris D., Taylor K.;
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RN Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapalate D., Elchner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL; AL031031; CAA19847.1;
 DR InterPro: IPR000653; DegT_Dnri_EryC1.
 DR Pfam: PF01041; DegT_Dnri_EryC1; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 225 AA; 24159 MW; CAEB7D9DDE7912C2 CRC64;
 Query Match 61.5%; Score 40; DB 16; Length 225;
 Best Local Similarity 77.8%; Pred. No. 86;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 4 DGGDGTVO 12
 Db 111111 11

DB 127 DGGGDTHTQ 135

RESULT 15

OSVTD9 PRELIMINARY: PRT: 418 AA.

AC OSVTD9: 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE CG6321 protein (LD15494P).

GN CG6321.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxId=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*."

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,

RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,

RA Nunoo J., Paclob J., Paragas V., Park S., Phouanavong S., Wan K.,

RA Xu C., Lewis S.E., Rubin G.M., Celiniker S.,

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AE003546; AAF50113.1; -

DR EMBL: AY061197; AAL28745.1; -

DR FlyBase: FBgn0036117; CG6321.

DR InterPro: IPR004839; AminoTransf1/2.

DR InterPro: IPR00408; Reg_chir-condens.

DR Pfam: PF00155; aminoTrans_1_2; 1.

DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.

SQ SEQUENCE 418 AA: 47413 MW; 257ACA959ECCABEE CRC64;

Query Match 61.5%; Score 40; DB 5; Length 418;

Best Local Similarity 70.0%; Pred. No. 17e+02;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 VFDGGDGTIV 11

DB 13 LFDGGDMVY 22

Search completed: June 16, 2003, 12:09:37

Job time: 6.02459 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:56:18 ; Search time 0.762295 Seconds
(without alignments)
652.918 Million cell updates/sec

Title: US-09-986-682B-2
Perfect score: 65
Sequence: 1 SVFDGSGDTVYQ 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	60.0	297	1	BMRU_BACSU
2	39	60.0	376	1	ACT1_LTP1
3	39	60.0	607	1	GDS1_HUMAN
4	38	58.5	418	1	LE21_THEMA
5	38	58.5	440	1	YA33_SCHPO
6	38	58.5	558	1	GDS1_BOVIN
7	37.5	57.7	473	1	SACB_BACSU
8	37	56.9	298	1	Y727_METUA
9	37	56.9	381	1	GKR2_ECOLI
10	37	56.9	382	1	GKR2_ECOLI
11	37	56.9	417	1	YAG1_YEAST
12	37	56.9	462	1	SAHH_RHIOA
13	37	56.9	463	1	SAHH_RHIOA
14	37	56.9	463	1	SAHH_RHIOA
15	37	56.9	466	1	SAHH_AGR15
16	37	56.9	466	1	SAHH_BROME
17	37	56.9	466	1	SAHH_RHIOA
18	37	56.9	466	1	SAHH_RHIOA
19	37	56.9	671	1	CHRA_BACSU
20	37	56.9	1018	1	CONT_HUMAN
21	37	56.9	1020	1	CONT_MOUSE
22	37	56.9	1021	1	CONT_MOUSE
23	37	56.9	1877	1	PKCS_MOUSE
24	37	56.9	1877	1	PKCS_MOUSE
25	36	55.4	115	1	KV2A_HUMAN
26	36	55.4	135	1	JANA_DROME
27	36	55.4	308	1	PYRB_METAC
28	36	55.4	371	1	GKR_METAC
29	36	55.4	371	1	GKR_METAC
30	36	55.4	381	1	GKR_METAC
31	36	55.4	394	1	GKR_METAC
32	36	55.4	421	1	GKR_METAC
33	36	55.4	496	1	ARAA_BACST

34	36	55.4	605	1	DNAX_AOUPY
35	36	55.4	747	1	TR15_ECOLI
36	36	55.4	1355	1	CA21_RANCA
37	36	55.4	4303	1	PKD1_HUMAN
38	35.5	54.6	1195	1	YEF1_YEAST
39	35	53.8	146	1	CUDA_LOCOMI
40	35	53.8	232	1	VHED_BPT7
41	35	53.8	232	1	VHED_BPT7
42	35	53.8	304	1	Y245_SYNY3
43	35	53.8	325	1	XYLB_BACOV
44	35	53.8	378	1	GKR_HAELN
45	35	53.8	380	1	GKR_BACHD

ALIGNMENTS

RESULT 1	ID	BMRU_BACSU	STANDARD:	PRT:	297 AA.
AC	P39074:				
DT	01-FEB-1995 (Rel. 31, Created)				
DT	01-FEB-1995 (Rel. 31, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Protein bmrU.				
GN	BMRU.				
OS	Bacillus subtilis.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
OX	NCBI_TaxID=1423;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168 / JH642;				
RX	MEDLINE=97124195; PubMed=8969508;				
RA	Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M., Kobayashi Y.;				
RT	"Systematic sequencing of the 283 kb 210 degrees-232 degrees region of the Bacillus subtilis genome containing the skin element and many sporulation genes.";				
RL	Microbiology 142:3103-3111(1996).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RX	MEDLINE=98044033; PubMed=9384377;				
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borchert S., Borst R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Broutelle S., Brusch C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Deniot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Ferrati E., Fougere D., Fritz C., Fujita M., Fujita Y., Funo S., Gallizi A., Galleron N., Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G., Guisepi G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karmata D., Kasahara Y., Kjaer-Blaugard M., Klein C., Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M., Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,				

RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Yamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambitt R., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RL Nature 390:249-256(1997).
 CC -1- SIMILARITY: TO E. COLI YEGS AND TO SYNECHOCYSTIS PCC 6803 SL10036.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L25604; AAB81538.1; -
 CC EMBL; D84432; BAA12602.1; -
 CC EMBL; Z99116; CAB14331.1; -
 CC Subtilist; BG10302; bmrV.
 CC InterPro; IPR005218; Cons_hypoth147.
 CC InterPro; IPR001206; DAGC.
 CC InterPro; IPR003622; DAG_kin_cat.
 CC Pfam; PF00781; DAGKC; 1.
 CC PRODOM; PD005043; DAG_kin_cat; 1.
 CC SMART; SM00046; DAGKC; 1.
 CC TIGRfams; TIGR00147; TIGR00147; 1.
 CC Complete proteome.
 CC KW SEQUENCE 297 AA; 32461 MW; 2EB5B9DC35F90CDD CRC64;
 SQ
 QY Query Match 60.0%; Score 39; DB 1; Length 297;
 DB Best Local Similarity 75.0%; Pred. No. 17;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 5 GGDGVYQ 12
 DB 66 GGDGTHQ 73
 RESULT 2
 ID ACT1_LYTP1 STANDARD; PRT; 376 AA.
 AC P53465;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Actin, cytoskeletal 1 (LPC1).
 OS *Lytechinus pictus* (Painted sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinozoa; Euechinozoa; Echinacea; Temnopleuroidea; Toxopneustidae;
 OC Lytechinus.
 CC NCBI_TaxID=7653;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95055791; PubMed=7966365;
 RA Fang H., Brandhorst B.P.;
 RT "Evolution of actin gene families of sea urchins."
 RL J. Mol. Evol. 39:347-356(1994).
 CC -1- FUNCTION: ACTINS ARE HIGHLY CONSERVED PROTEINS THAT ARE INVOLVED
 CC IN VARIOUS TYPES OF CELL MOTILITY AND ARE UBICUITOUSLY EXPRESSED
 CC IN ALL EUKARYOTIC CELLS.
 CC -1- SUBCELLULAR LOCATION: CYTOSKELETAL.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN SEVERAL DIFFERENT SPATIAL
 CC TERRITORIES OF THE EMBRYO.
 CC -1- SIMILARITY: BELONGS TO THE ACTIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U09651; AAA53363.1; -
 CC HSSP; P02570; ZBT.
 CC InterPro; IPR004001; Actin.
 CC InterPro; IPR004000; Actin_like.
 CC Pfam; PF00022; actin; 1.
 CC PRINTS; PR00190; ACTIN.
 CC SMART; SM00268; ACTIN; 1.
 CC PROSITE; PS00406; ACTINS_1; FALSE_NEG.
 CC PROSITE; PS00432; ACTINS_2; 1.
 CC PROSITE; PS01132; ACTINS_ACT LIKE; 1.
 CC KW Structural protein; Multigene family; Acetylation.
 CC FT PROPEP 1 2 REMOVED IN MATURE FORM (BY SIMILARITY).
 CC FT CHAIN 3 376 ACTIN, CYTOSKELETAL 1.
 CC FT MOD_RES 3 376 ACETYLTATION (BY SIMILARITY).
 CC FT SEQUENCE 376 AA; 41842 MW; 34F9C757A1F65C90 CRC64;
 SQ
 QY Query Match 60.0%; Score 39; DB 1; Length 376;
 DB Best Local Similarity 70.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 VPDGDDGVY 11
 DB 153 VPDGDDGVY 162
 RESULT 3
 ID GDS1_HUMAN STANDARD; PRT; 607 AA.
 AC P52306; Q9NZ48; Q9NYM2;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Rapi GTPase-GDP dissociation stimulator 1 (SMG p21 stimulatory GDP/GTP
 DE exchange protein) (SMG GDS protein) (Exchange factor smgGDS).
 GN RAP1GDS1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=92195558; PubMed=1549351;
 RA Kikuchi A., Kibuchi K., Hoti Y., Nonaka H., Sakoda T.,
 RA Kawamura M., Mizuno T., Takai Y.;
 RT "Molecular cloning of the human cDNA for a stimulatory GDP/GTP
 RT exchange protein for c-Ki-ras p21 and smg p21."
 RL Oncogene 7:289-293(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Vikis H.G., Stewart S., Guan K.;
 RT "SmgGDS as an N17ras binding protein."
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Peripheral blood;
 RA Hussey D.J., Albanese N.O., Dobrovic A.;
 RT "The major isoform of RAP1GDS1 mRNA in peripheral blood mononuclear
 RT cells encodes an isoform of smgGDS with 12 armadillo repeats."
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: STIMULATES GDP/GTP EXCHANGE REACTION OF A GROUP OF SMALL
 CC GTP-BINDING PROTEINS (G PROTEINS) INCLUDING RAP1A/RAP1B, RHOA,
 CC RHOB AND C-KI-RAS P21, BY STIMULATING THE DISSOCIATION OF GDP FROM
 CC AND THE SUBSEQUENT BINDING OF GTP TO EACH SMALL G PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
 CC produced by alternative splicing.
 CC -1- SIMILARITY: CONTAINS 5 ARM REPEATS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: X63465; CAA45067.1; -
 DR EMBL: AF215923; AAF32290.1; -
 DR EMBL: AF237413; AAF43211.1; -
 DR Genew: HGNC:9859; RAP1GDSL.
 DR MIM: 179502; -
 DR InterPro: IPR000225; Armadillo.
 DR Pfam: PF00514; Armadillo_seg; 5.
 DR SMART: SM00185; ARM; 4.
 DR PROSITE: PS50176; ARM_REPEAT; 2.
 KW Cytase activation; Repeat; Alternative splicing.
 FT REPEAT 77 118 ARM 1.
 FT REPEAT 169 211 ARM 2.
 FT REPEAT 346 390 ARM 3.
 FT REPEAT 391 431 ARM 4.
 FT REPEAT 479 519 ARM 5.
 FT VARSPLIC 122 170 MISSING (IN ISOFORM 2).
 FT CONFLICT 376 376 T -> P (IN REF. 3).
 FT CONFLICT 394 394 D -> N (IN REF. 3).
 FT CONFLICT 593 593 R -> A (IN REF. 3).
 SQ SEQUENCE 607 AA: 66402 MW: 074F9197BA98782P CRC64;

Query Match 60.0%; Score 39; DB 1; Length 607;
 Best Local Similarity 54.5%; Pred. No. 37;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Oy 2 VFDDGDTGVYQ 12
 :1:111111:
 Db 309 LFEGCKGSVQ 319

RESULT 4
 LE21_THEME STANDARD; PRT: 418 AA.
 AC 09WYC;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 3-Isopropylmalate dehydratase large subunit 1 (EC 4.2.1.33)
 DE (Isopropylmalate isomerase 1) (Alpha-IPM isomerase 1) (IPMI 1).
 GN LEUC1 OR TM0291.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales;
 OC Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX NEJLON=99287316; PubMed=10360571;
 RA NEJLON K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
 RA Hatt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Uitterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima";
 RL Nature 399:323-329(1999).

CC -1- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate
 CC and 3-isopropylmalate, via the formation of 2-isopropylmalate.
 CC -1- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmalate +
 CC H₂O.
 CC -1- CATALYTIC ACTIVITY: 2-isopropylmalate + H₂O = 2-
 CC isopropylmalate.
 CC -1- PATHWAY: Leucine biosynthesis; second step
 CC -1- SUBUNIT: Heterodimer of leuc and leud (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY. LEUC 2

CC SUBFAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: AE001711; AAD35379.1; -
 DR HSSP: P20004; IAMJ.
 DR TIGR: TM0291; -
 DR InterPro: IPR001030; Aconitase_N.
 DR Pfam: PF00330; aconitase; 1.
 DR PRINTS: PR00415; ACONITASE.
 DR Prodom: PD000511; Aconitase_N; 1.
 DR PROSITE: PS00450; ACONITASE_1; 1.
 DR PROSITE: PS01244; ACONITASE_2; 1.
 KW Leucine biosynthesis; Lyase; Iron-sulfur; 4Fe-4S; Complete proteome.
 FT METAL 298 298 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 358 358 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 361 361 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 418 AA: 45447 MW: 1BD41B4BFA825BF CRC64;

Query Match 58.5%; Score 38; DB 1; Length 418;
 Best Local Similarity 63.6%; Pred. No. 36;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 2 VFDDGDTGVYQ 12
 :111111:
 Db 91 VFDDGDTGVYQ 101

RESULT 5
 YA33_SCHPO STANDARD; PRT: 440 AA.
 AC 009710;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C18B11.03c in chromosome I.
 GN SPAC18B11.03c.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Roben J., Glynnoprez B.,
 RA Welljens I., Vansteelandt E., Rieger M., Scheffer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leinrich H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambolt R., Punelle B.,
 RA Coffeau A., Cadieu E., Dreano S., Gloux S., Lélaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA	Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA	Cerniichi L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA	Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.,
RL	"The genome sequence of <i>Schizosaccharomyces pombe</i> ."
RL	Nature 413:871-880(2002).
CC	- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC	the European Bioinformatics Institute. There are no restrictions on its use
CC	by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce).
CC	or send an email to license@isb-sib.ch).
DR	EMBL: Z50728; CAA90588.1; -
KW	Hypothetical protein: Transmembrane.
FT	TRANSMEM 21 41 POTENTIAL.
FT	TRANSMEM 234 254 POTENTIAL.
FT	TRANSMEM 390 410 POTENTIAL.
SO	SEQUENCE 440 AA; 49579 MW; CACB9F3EDDFE89AC CRC64;
Query Match	58.5%; Score 38; DB 1; Length 440;
Best Local Similarity	54.5%; Pred. NO. 39;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0	
QY	1 SVFDGDCGVY 11
	1: 11::
Db	126 SLFDGSGGLF 136
RESULT 6	
GDS1_BOVIN	STANDARD: PRT; 558 AA.
AC	Q04173; GPTS36;
DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Rapl GTPase-GDP dissociation stimulator 1 (SMG p21 stimulatory GDP/GTP
DE	exchange protein) (SMG GDS protein) (Exchange factor smgGDS).
GN	RAP1GDS1.
OS	Bos taurus (Bovine).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC	Bovidae; Bovinae; Bos.
OX	NCBI_TaxID:9913;
RN	[1]
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC	TISSUE=Brain;
RC	MEDLINE=91203910; PubMed=1901951;
RA	Kaibuchi K., Mizuno T., Fujio H., Yamamoto T., Kishi K.,
RA	Fukumoto Y., Hori Y., Takai Y.,
RT	"Molecular cloning of the cDNA for stimulatory GDP/GTP exchange
RT	protein for smg p21s (ras p21-like small GTP-binding proteins) and
RT	characterization of stimulatory GDP/GTP exchange protein."
RL	Mol. Cell. Biol. 11:2873-2880(1991).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Brain;
RC	MEDLINE=92356593; PubMed=1495270;
RA	Yamamoto T.;
RT	"The stimulatory GDP/GTP exchange protein for ras p21-related small
RT	GTP-binding proteins."
RL	Kobe J. Med. Sci. 38:37-56(1992).
RN	[3]
RP	SEQUENCE OF 350-356.
RC	TISSUE=Brain;
RC	MEDLINE=90375538; PubMed=2118909;
RA	Yamamoto T., Kaibuchi K., Mizuno T., Hirogoshi M., Shitake H.,
RA	Takai Y.;
RT	"Purification and characterization from bovine brain cytosol of
RT	proteins that regulate the GDP/GTP exchange reaction of smg p21s, ras
RT	p21-like GTP-binding proteins."

```

RL J. Biol. Chem. 265:16626-16634(1990).
CC -i- FUNCTION: STIMULATES GDP/GTP EXCHANGE REACTION OF A GROUP OF SMALL
CC GTP-BINDING PROTEINS (G PROTEINS) INCLUDING RAP1A/RAP1B, RHOA,
CC RHOB AND C-K1-RAS P21, BY STIMULATING THE DISSOCIATION OF GDP FROM
CC AND THE SUBSEQUENT BINDING OF GTP TO EACH SMALL G PROTEIN.
CC -i- TISSUE SPECIFICITY: BRAIN.
CC -i- PTM: THE N-TERMINUS IS BLOCKED.
CC -i- SIMILARITY: CONTAINS 3 ARM REPEATS.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M63325; AAA1876..1;
CC InterPro: IPR000225; Armadillo.
CC Pfam: PF00514; Armadillo_seg; 5.
CC SMART: SM00185; ARM_4.
CC PROSITE: PS50176; ARM_REPEAT; 2.
CC KW GTPase activation; Repeat.
CC FT REPEAT 79 118 ARM 1.
CC FT REPEAT 121 162 ARM 2.
CC FT REPEAT 298 341 ARM 3.
CC FT CONFLICT 262 325
CC ECGKGNVFRVRLSWTPSNHQLQALALANFARNDNCI
CC HAVMDIVTEKLMDDLDRHVEGN -> VTYGHAALSCRLN
CC AIPVVKAKKMLSGAVTEAVLKFELKSEMPVQFKLTGLRL
CC IDAQMOKLF (IN REF. 2).
CC AAEOLGKRVKRYERVWCEAKDHAGVGESNRLLSALIRH
CC SKSRKVITYIQSGGIRKLTYMA -> TSEYIYMNELVLA
CC LALLIALELGTAEKDLSEAOQLVQILHRLAEDRSAPETRYN
CC SMVLICAAE (IN REF. 2).
CC
CC SQ SEQUENCE 558 AA; 61066 MW; C04C5F435B9B297F CRC64;
CC
CC Query Match 58.5%; Score 38; DB 1; Length 558;
CC Best Local Similarity 54.5%; Pred. NO. 50;
CC Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0.
CC
CC QY 2 VFDDGGDGTIVYQ 12
CC :|:|:|:|:|
CC DB 260 LPEGGKGWVFQ 270
CC
CC RESULT 7
CC SACB_BACSU STANDARD: PRT; 473 AA.
CC ID SACB_BACSU
CC AC P05655; P70984;
CC DT 01-NOV-1988 (Rel. 09, Created)
CC DT 01-NOV-1988 (Rel. 09, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Levanucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
CC transferase) (Sucrose 6-fructosyl transferase).
CC GN SABC.
CC OS Bacillus subtilis.
CC OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CC OX NCBI_TaxID=1423;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=168;
CC RA MEDLINE=85295507; PubMed=2993618;
CC RA Steinhilber M., Le Cog D., Aymeric S., Gonzy-Treboul G., Gay P.;
CC "The DNA sequence of the gene for the secreted Bacillus subtilis
CC enzyme levanucrase and its genetic control sites.";
CC RL Mol. Gen. Genet. 200:220-228(1985).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=168;
CC RA Denizot F.;
CC RN Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC [3]
CC RP SEQUENCE FROM N.A.

```


CC STRAIN-168;
 RX MEDLINE=96044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertoro M.G., Bessieres P., Bolotin A., Borcherdt S.,
 RA Borriss R., Boutsier L., Brans A., Braun M., Briganti S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codan J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Deutz F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entlan K.D., Erington J., Fabret C., Ferrari E., Foulger D.,
 RA Filtz C., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henauf A.,
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapides A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogilawa A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Priesen E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serrif P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Weltenegeger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zandich A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RT Nature 390:249-256(1997).
 RL [4]
 RN SEQUENCE OF 1-62 FROM N.A.
 RP STRAIN-168 / PY79;
 RC MEDLINE=84178454; PubMed=6424671;
 RX Fouet A., Arnaud M., Klier A., Rapoport G.;
 RA "Characterization of the precursor form of the exocellular
 RT levansucrase from *Bacillus subtilis*.";
 RL Biochem. Biophys. Res. Commun. 119:795-800(1984).
 RN [5]
 RP SEQUENCE OF 1-68 FROM N.A.
 RX MEDLINE=87008406; PubMed=2428811;
 RA Shimotsu H., Hener D.J.;
 RT "Modulation of *Bacillus subtilis* levansucrase gene expression by
 RT sucrose and regulation of the steady-state mRNA level by *sacY* and
 RT *sacQ* genes.";
 RL J. Bacteriol. 168:380-388(1986).
 CC -1- CATALYTIC ACTIVITY: Sucrose + [(2,6)-beta-D-fructosyl](N) =
 CC glucose + [(2,6)-beta-D-fructosyl](N+1).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 CC EMBL: M14202; AAA22725.1; -;
 DR EMBL: 294043; CAB08015.1; -;
 DR EMBL: 299121; CAB15450.1; -;
 DR EMBL: K01987; AAA22724.1; -;
 DR EMBL: X02730; CAA26513.1; -;
 DR PIR: A25040; A25040.
 DR PIR: S07309; S07309.
 DR Subtilist; BG10388; sacB.
 DR InterPro: IPR003469; GLYCO_hydro_68.
 DR Pfam: PF02435; Glycosyltransferase; Signal; Complete proteome.
 KW Transferase; Glycosyltransferase; Signal; Complete proteome.

FT SIGNAL 1 29
 FT CHAIN 30 473
 FT CONFLICT 12 12 V -> I (IN REF. 4).
 SQ SEQUENCE 473 AA: 52971 MW: 38F2F571B41D5B0 CRC64;
 Query Match 57.7%; Score 37.5; DB 1; Length 473;
 Best Local Similarity 66.7%; Pred. No. 51;
 Matches 8; Conservative 1; Mismatches 2; Indels 1; Gaps 1;
 QY 1 SWPDGDDGTYYQ 12
 Db 216 SLFD-GDGRTYYQ 226
 RESULT 8
 ID Y727_METJA STANDARD; PRT; 298 AA.
 AC 058137;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0727.
 GN MJ0727.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_Taxid=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.A., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, *Methanococcus*
 RT *jannaschii*.";
 RL Science 273:1058-1073(1996).
 CC -1- SIMILARITY: SOME. TO COENZYME F420 HYDROGENASE ALPHA SUBUNIT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 CC EMBL: U67519; AAB98723.1; -;
 DR TIGR: M07027;
 DR InterPro: IPR001501; NI_hdt.
 DR Pfam: PF00374; NifHase_Hases; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 298 AA: 33519 MW: 4DAD534096A680D CRC64;
 Query Match 56.9%; Score 37; DB 1; Length 298;
 Best Local Similarity 54.5%; Pred. No. 37;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 SWPDGDDGTYY 11
 Db 241 AVIEGEGIVY 251
 RESULT 9
 ID GRK2_ECOLI STANDARD; PRT; 381 AA.
 AC P23524;
 DT 01-NOV-1991 (Rel. 20, Created)

```

DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glycerate kinase 2 (EC 2.7.1.31).
GN GAK OR B3124.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91154140; PubMed=1705543;
RA Komine Y., Inokuchi H.;
RT "Precise mapping of the rnpB gene encoding the RNA component of RNase
RL P. in Escherichia coli K-12."
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1233-1238(1997).
CC -i- CATALYTIC ACTIVITY: ATP + (R)-glycerate = ADP + 3-phospho-(R)-
CC glycerate.
CC -i- SIMILARITY: BELONGS TO THE GLYCERATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D90212; BAA14239.1; ALT_INIT.
CC EMBL: U18997; AAA57927.1; ALT_INIT.
CC EMBL: AE000394; AAC76158.1; ALT_INIT.
CC PIR: J00614; J00614.
CC Ecogen: EG11175; gark.
CC InterPro: IPR004381; Cons_hypoth45.
CC InterPro: IPR003747; Glycerate_kinase.
CC Pfam: PF02595; DUF168.1.
CC TIGRFAMs: TIGR00045; Cons_hypoth45; 1.
CC Transfaser: Kinase; Complete proteome.
CC SEQUENCE 381 AA; 39104 MW; 0A048E2E5F9FE32 CRC64;

```

```

RX MEDLINE=96093926; PubMed=7584049;
RA Yoshida K.-I., Seki S., Fujimura M., Miwa Y., Fujita Y.;
RT "Cloning and sequencing of a 36-kb region of the Bacillus subtilis
RL genome between the gnt and iol operons."
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunz F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Boloitin A., Borcher S.,
RA Bortiss R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koettler P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portelle D., Porvillik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle R., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeder R., Scifone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takauchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarelli A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Wellenreger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein E.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RL subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 1-13 FROM N.A.
RX MEDLINE=87008613; PubMed=3020045;
RA Fujita Y., Fujita T., Miwa Y., Nishashi J., Aratani Y.;
RT "Organization and transcription of the gluconate operon, gnt, of
RT Bacillus subtilis."
RL J. Biol. Chem. 261:13744-13753(1986).
CC -i- CATALYTIC ACTIVITY: ATP + (R)-glycerate = ADP + 3-phospho-(R)-
CC glycerate.
CC -i- SIMILARITY: BELONGS TO THE GLYCERATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AB005554; BAA21580.1; -
CC EMBL: Z99124; CAB16041.1; -
CC EMBL: X02584; -; NOT_ANNOTATED_CDS.
CC EMBL: X03510; -; NOT_ANNOTATED_CDS.
CC Subtilast: BG1103; glxK.
CC InterPro: IPR004381; Cons_hypoth45.
CC InterPro: IPR003747; Glycerate_kinase.
CC Pfam: PF02595; DUF168.1.
CC TIGRFAMs: TIGR00045; Cons_hypoth45; 1.
CC Transfaser: Kinase; Complete proteome.
CC SEQUENCE 382 AA; 39398 MW; 9242CB53C0DA5EDB CRC64;

```

```

Query Match 56.9%; Score 37; DB 1; Length 382;
Best Local Similarity 77.8%; Pred. No. 49;

```

```

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 2 VFDGDDCTV 10
   |1111111|
Db 41 VADGCESTV 49

RESULT 11
YAGI_YEAST
ID YAGI_YEAST STANDARD: PRT: 417 AA.
AC P39713;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical zinc-type alcohol dehydrogenase-like protein in GDH3-CNE1
DE Intergenic region.
GN VAL061W OR FUN50.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kueback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae."
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC -1- COFACTOR: ZINC (Potential).
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U12980; AAC04973.1; -.
DR SGD: S0000057; YAL061W.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR Pfam: PF00107; adh_zinc.1.
DR PROSITE: PS00059; ADH_ZINC: 1.
KW Hypothetical protein; Oxidoreductase; Zinc.
FT METAL 39 39 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 64 64 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 120 120 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 123 123 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 131 131 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 173 173 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 417 AA; 46098 MW; D921CEDBE61D0151 CRC64;

Query Match 56.9%; Score 37; DB 1; Length 417;
Best Local Similarity 63.6%; Pred. No. 54;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 SVFDGDDGTV 11
   |1111111|
Db 254 SIADGDDGDFY 264

RESULT 12
SAHH_ROSDE STANDARD: PRT: 462 AA.
AC Q92NA5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adenosylhomocysteine (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE hydrolase) (AdoHcyase).
GN AHcy.
OS Roseobacter denitrificans (Erythrobacter sp. (strain OCH 114)).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Roseobacter.
OX NCBI_TaxID=2434;
RN [1]
RP SEQUENCE FROM N.A.
RA Nishimura K., Shimada H., Shimen T., Odayashi T., Masuda T., Ohta H.,
RA Takamiya K.;
RT "Photosynthetic regulatory gene cluster in an aerobic photosynthetic
RT bacterium, Roseobacter denitrificans."
RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC + L-homocysteine.
CC -1- COFACTOR: NAD (By similarity).
CC -1- PATHWAY: Activated methyl cycle.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB020211; BAA34645.1; -.
DR HSSP: P10760; 1B3R.
DR InterPro: IPR000043; Ado_hcyase.
DR InterPro: IPR000205; NAD_binding.
DR Pfam: PF00670; AdoHcyase; 1.
DR TIGRFAMs: TIGR00936; ahcy; 1.
DR PROSITE: PS00738; ADOHCYASE_1; 1.
DR PROSITE: PS00739; ADOHCYASE_2; 1.
KW Hydrolase; NAD: One-carbon metabolism.
NP BIND 246 277 NAD (POTENTIAL).
SQ SEQUENCE 462 AA; 50484 MW; 8D8D3A6F3F896FBE CRC64;

Query Match 56.9%; Score 37; DB 1; Length 462;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 DGDGCTV 11
   |1111111|
Db 128 DGDGATV 135

RESULT 13
SAHH_RHOCA STANDARD: PRT: 463 AA.
AC P28183;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adenosylhomocysteine (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE hydrolase) (AdoHcyase).
GN AHcy.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SBI003 / St Louis;
RX MEDLINE=92335291; PubMed=1631127;
RA Sganga M.W., Aksent R.R., Cantoni G.L., Bauer C.E.;
RT "Mutational and nucleotide sequence analysis of S-adenosyl-L-
RT homocysteine hydrolase from Rhodobacter capsulatus."
RL Proc. Natl. Acad. Sci. U.S.A. 89:6328-6332(1992).

```

```

RN [2]
RP SEQUENCE OF 1-13 FROM N.A.
RC STRAIN-SBI003 / St Louis;
RX MEDLINE=94110241; PubMed=8282711;
RA Buggy J.J., Sganga M.W., Bauer C.E.;
RT "Nucleotide sequence and characterization of the Rhodobacter
RT capsulatus hvrB gene: HvrB is an activator of S-adenosyl-L-
RT homocysteine hydrolase expression and is a member of the LysR
RT family.";
RL J. Bacteriol. 176:61-69(1994).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC + L-homocysteine.
CC -1- COFACTOR: NAD.
CC -1- PATHWAY: Activated methyl cycle.
CC -1- SUBUNIT: Homotetramer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE ADENOSYLMHOMOCYSTEINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M80630; AAA26094.1; -
CC EMBL: L23836; AAA53540.1; -
CC HSSP: P10760; 1B3R.
CC InterPro: IPR000043; Ado_hcyase.
CC Pfam: PF00670; AdoHcyase; 1.
CC TIGRFAMs: TIGR00936; ancy; 1.
CC PROSITE: PS00738; ADOHCYASE_1; 1.
CC PROSITE: PS00739; ADOHCYASE_2; 1.
CC Hydrolyase: NAD: One-carbon metabolism.
CC NP_BIND: 247 278 NAD (POTENTIAL).
CC SEQUENCE 463 AA; 50580 MW; E4EB19B320420B53 CRC64;
SQ
Query Match 56.9%; Score 37; DB 1; Length 463;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 4 DGGDCTVY 11
Db 1111 1:1
128 DGGDATTY 135
RESULT 14
SAHH_RHOSH STANDARD; PRT; 463 AA.
ID SAHH_RHOSH
AC O50562;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adenosylhomocysteinase (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE hydrolase) (AdoHcyase).
GN AHcy.
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97354111; PubMed=9210332;
RA Mizoguchi H., Masuda T., Nishimura K., Shimada H., Ohta H., Shioi Y.,
RA Takamiya K.;
RT "Nucleotide sequence and transcriptional analysis of the flanking
RT region of the gene (spb) for the trans-acting factor that controls
RT light-mediated expression of the puf operon in Rhodobacter
RT sphaeroides.";
RL Plant Cell Physiol. 38:558-567(1997).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC + L-homocysteine.

```

```

CC -1- COFACTOR: NAD (By similarity).
CC -1- PATHWAY: Activated methyl cycle.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE ADENOSYLMHOMOCYSTEINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U76671; AAB88245.1; -
CC HSSP: P10760; 1B3R.
CC InterPro: IPR000043; Ado_hcyase.
CC Pfam: PF00670; AdoHcyase; 1.
CC TIGRFAMs: TIGR00936; ancy; 1.
CC PROSITE: PS00738; ADOHCYASE_1; 1.
CC PROSITE: PS00739; ADOHCYASE_2; 1.
CC Hydrolyase: NAD: One-carbon metabolism.
CC NP_BIND: 247 278 NAD (POTENTIAL).
CC SEQUENCE 463 AA; 50617 MW; ACDBE9D0D0CCBBA CRC64;
SQ
Query Match 56.9%; Score 37; DB 1; Length 463;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 4 DGGDCTVY 11
Db 1111 1:1
128 DGGDATTY 135
RESULT 15
SAHH_AGR5 STANDARD; PRT; 466 AA.
ID SAHH_AGR5
AC O80099;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adenosylhomocysteinase (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE hydrolase) (AdoHcyase).
GN AHcy OR ARU0029 OR AGR_C_46.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gatlung S., Blanchard M.,
RA Quicollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughly D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Curson J., Lomo C., Sear C., Strub G.,
RA Cléo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent

```

```
. RT Agrobacterium tumefaciens C58.
RL Science 294:2323-2328(2001).
CC -I- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC + L-homocysteine.
CC -I- COFACTOR: NAD (By similarity).
CC -I- PATHWAY: Activated methyl cycle.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008977; AAA1060.1; -
DR EMBL; AE007946; AAK8583.1; -
DR PROSITE; PS00738; ADOHCYASE_1; 1.
DR PROSITE; PS00739; ADOHCYASE_2; 1.
KW Hydrolase; NAD; One-carbon metabolism; Complete proteome.
FT NP BIND 250 281 NAD (PORENTIAL).
SQ SEQUENCE 466 AA; 51159 MW; 9A124B304419F26 CRC64;

Query Match 56.9%; Score 37; DB 1; Length 466;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 DGGDCTV 11
   1111 1:1
Db 132 DGGDATMY 139
```

Search completed: June 16, 2003, 12:07:25
Job time : 1.7623 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 16, 2003, 12:03:38 ; Search time 1.47541 Seconds
(without alignments)
781.893 Million cell updates/sec

Title: US-09-986-682B-2
Perfect score: 65
Sequence: 1 SVFDGNGTIVQ 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	90.8	489	2 F97118	levansucrase [impo
2	43	66.2	330	2 D87393	hypothetical prote
3	42	64.6	767	2 T30018	hypothetical prote
4	41	63.1	295	2 D83734	hypothetical prote
5	41	63.1	380	2 T35953	conserved hypothe
6	40	61.5	225	2 T35679	hypothetical prote
7	40	61.5	469	2 T35670	hypothetical prote
8	39	60.0	108	2 B86321	hypothetical prote
9	39	60.0	297	2 F69595	multidrug resistan
10	39	60.0	523	2 T11317	cytochrome-c oxida
11	39	60.0	518	2 I37456	smg GDS - human
12	39	60.0	655	2 JE0358	raib binding prote
13	38.5	59.2	647	2 A84265	hypothetical prote
14	38.5	59.2	815	2 T40524	hypothetical prote
15	38	58.5	146	2 C86187	hypothetical prote
16	38	58.5	382	2 T49762	hypothetical prote
17	38	58.5	388	2 JY0308	cellulase (EC 3.2.
18	38	58.5	398	2 H70393	hemolysin - Aquife
19	38	58.5	418	2 C72394	hypothetical prote
20	38	58.5	440	2 S58302	hypothetical prote
21	38	58.5	558	2 A38795	stimulatory GDP/GT
22	38	58.5	558	2 T48499	receptor-like prot
23	37.5	57.7	473	2 A25040	levansucrase (EC 2
24	37	56.9	82	2 S65779	gamma-thionin prec
25	37	56.9	249	2 G84324	hypothetical prote
26	37	56.9	298	2 G64390	coenzyme F420 hydr
27	37	56.9	311	2 F83871	multidrug resistan
28	37	56.9	342	2 G72567	hypothetical prote
29	37	56.9	377	2 E82402	conserved hypothet

30	37	56.9	379	2 C75167	hypothetical prote
31	37	56.9	381	2 AD0897	conserved hypothet
32	37	56.9	382	2 A70071	conserved hypothet
33	37	56.9	387	2 B85974	hypothetical prote
34	37	56.9	408	2 J00614	hypothetical prote
35	37	56.9	408	2 B91129	ypad protein - Bsc
36	37	56.9	417	2 S51961	hypothetical prote
37	37	56.9	462	1 A46035	FUN50 protein - ye
38	37	56.9	466	2 D97362	adenosylhomocystei
39	37	56.9	466	2 AP2580	adenosylhomocystei
40	37	56.9	481	2 AG3505	S-adenosylhomocyst
41	37	56.9	620	2 T30765	adenosylhomocystei
42	37	56.9	629	2 H86383	hypothetical prote
43	37	56.9	671	1 ORSCN	probable wall-asso
44	37	56.9	747	2 A71440	two-component sens
45	37	56.9	801	2 G82302	hypothetical prote
					probable cellobios

ALIGNMENTS

RESULT 1
F97118
levansucrase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: F97118
R.Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.
J. Daly, M.D.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97118
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-489 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79737.1; PID:q15024742; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1772

Query Match
Best Local Similarity 90.8%; Score 59; DB 2; Length 489;
Matches 10; Conservative 83.3%; Pred. No. 0.031;
Mismatch 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVFDGNGTIVQ 12
|||||:|:|:
DB 229 SVFDGNGTIVQ 240

RESULT 2
D87393
hypothetical protein CC1160 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: D87393
R.Nietman, W.C.; Feldblym, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Dirkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Fomicheva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: D87393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-330 <STO>
A:Cross-references: GB:AE005673; NID:q13422480; PIDN:AAK23144.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1160

Query Match
Best Local Similarity 66.2%; Score 43; DB 2; Length 330;
Matches 8; Conservative 88.9%; Pred. No. 10;
Mismatch 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VFDDGDTGV 10
| | | | |
Db. 80 VIDGGDTGV 88

RESULT 3

150018
hypothetical protein F38B9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T30018
R:Mu, X.; Gattung, S.
Submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F38B9.
A:Reference number: 220722
A:Accession: T30018
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-767 <WUX>
A:Cross-references: EMBL:U46668; PIDN:AAA93348.1; CESP:F38B9.2
C:Genetics:

A:Gene: CESP:F38B9.2
A:Introns: 13/1; 37/3; 62/1; 89/1; 127/1; 164/2; 182/1; 235/3; 269/3; 321/2; 360/1; 455/

Query Match 64.6%; Score 42; DB 2; Length 767;
Best Local Similarity 77.8%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVFDGDTGV 9
| | | | |
Db 592 SYDDGDTGV 600

RESULT 4

DB3734
hypothetical protein BH0676 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: DB3734
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirz
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: DB3734
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-295 <STO>
A:Cross-references: GB:AP001509; GB:BA000004; NID:q10173176; PIDN:BAH04395.1; GSPDB:GNOC
C:Genetics:
A:Experimental source: strain C-125
A:Gene: BH0676

Query Match 63.1%; Score 41; DB 2; Length 295;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 GGDGTGVQ 12
| | | | |
Db 65 GGDGTIVE 72

RESULT 5

T35953
conserved hypothetical protein SC9C7.02 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T35953
R:Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
Submitted to the EMBL Data Library, January 1999
A:Reference number: Z21551
A:Accession: T35953
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-380 <SEE>
A:Cross-references: EMBL:AL035161; PIDN:CAA22714.1; GSPDB:GN00070; SCOEDB:SC9C7.02
C:Genetics:
A:Gene: SCOEDB:SC9C7.02
C:Superfamily: yhad protein

Query Match 63.1%; Score 41; DB 2; Length 380;
Best Local Similarity 88.9%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VFDDGDTGV 10
| | | | |
Db 45 VADGGDTGV 53

RESULT 6

T35679
hypothetical protein SC7C7.01 SC7C7.01 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35679
R:Harris, D.; Taylor, K.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
Submitted to the EMBL Data Library, July 1998
A:Reference number: Z21587
A:Accession: T35679
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-225 <HAR>
A:Cross-references: EMBL:AL031031; PIDN:CAA19847.1; GSPDB:GN00070; SCOEDB:SC7C7.01
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC7C7.01

Query Match 61.5%; Score 40; DB 2; Length 225;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 DGGDTGVQ 12
| | | | |
Db 127 DGGDTHTQ 135

RESULT 7

T35670
hypothetical protein SC7B7.02 SC7B7.02 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35670
R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
Submitted to the EMBL Data Library, November 1997
A:Reference number: Z21586
A:Accession: T35670
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-469 <MUR>
A:Cross-references: EMBL:AL009199; PIDN:CAA15782.1; GSPDB:GN00070; SCOEDB:SC7B7.02
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC7B7.02

Query Match 61.5%; Score 40; DB 2; Length 469;
Best Local Similarity 63.6%; Pred. No. 48;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SVFDGDTGV 11
| | | | |
Db 65 AIFDGGDTGV 75

RESULT 8

B86321

hypothetical protein F6A14.16 [Imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
 C:Accession: B86321
 R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 Ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: B86321
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-108 <STO>
 A:Cross-references: GB:AE005172; NID:g6730711; PIDN:AAF27106.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 60.0%; Score 39; DB 2; Length 108;
 Best Local Similarity 87.5%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 DGGDGTYY 11
 |||||
 DB 59 DGGDGTGY 66

RESULT 9
 F69595

multidrug resistance protein cotranscribed with bmr bmr - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
 C:Accession: F69595

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter,
 C.; Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho,
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Funai, S.; Galizzi, A.; Gallier,
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel,
 Y.; M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
 A:Authors: Schleich, S.; Schroeter, R.; Scifone, F.; Sekiguchi, J.; Sekowska, A.; Serot,
 akuchl, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zunshtein, E.; Yoshikawa, H.; Danculin, A.
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:96044033; PMID:9384377

A:Accession: F69595
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-297 <KUN>
 A:Cross-references: GB:299116; GB:AL009126; NID:g2634723; PIDN:CA814331.1; PID:ell185668;
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: bmr

Query Match 60.0%; Score 39; DB 2; Length 297;
 Best Local Similarity 75.0%; Pred. No. 44;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 5 GGDGTYYO 12
 |||||
 DB 66 GGDGTTHQ 73

RESULT 10

T11317
 cytochrome-c oxidase (EC 1.9.3.1) chain I - Pedinomonas minor mitochondrion
 C:Species: mitochondrion Pedinomonas minor
 C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
 C:Accession: T11317
 R:Turner, M.; Lemieux, C.; Burger, G.; Lang, B.F.; Otis, C.; Plante, I.; Gray, M.W.
 submitted to the EMBL Data Library, December 1998
 A:Description: The complete mitochondrial DNA sequences of Nephrolepis olivacea and
 A:Reference number: Z17261

A:Accession: T11317
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-523 <TUR>
 A:Cross-references: EMBL:AF116775; NID:g4378766; PID:g4378771; PIDN:AA019669.1
 C:Genetics:

A:Gene: cox1
 A:Genome: mitochondrion
 A:Genetic code: SGC3
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C:Keywords: copper; electron transfer; heme; iron; magnesium; membrane-associated com
 P.12-459/Domain: cytochrome-c oxidase chain I homology <CO1>
 F:64,380/Binding site: heme a iron (His) (axial ligands) #status predicted
 F:243,292/Binding site: copper (His) #status predicted
 F:243,247/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F:247/Binding site: oxygen (Tyr) #status predicted
 F:370/Binding site: magnesium (His) (shared with chain II) #status predicted
 F:378/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 60.0%; Score 39; DB 2; Length 523;
 Best Local Similarity 58.3%; Pred. No. 78;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 SVFDGGDGTYYO 12
 ||:||||
 DB 224 SVFGGDPVLYO 235

RESULT 11
 137456

smg GDS - human
 C:Species: Homo sapiens (man)
 C>Date: 04-Oct-1996 #sequence_revision 04-Oct-1996 #text_change 05-Nov-1999
 C:Accession: 137456; S22144
 R:Kikuchi, A.; Kibuchi, K.; Horii, Y.; Nonaka, H.; Sakoda, T.; Kawamura, M.; Mizuno,
 Oncogene 7, 289-293, 1992
 A>Title: Molecular cloning of the human cDNA for a stimulatory GDP/GTP exchange prote
 A:Reference number: 137456; MUID:92195658; PMID:1549351

A:Accession: 137456
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-558 <RES>
 A:Cross-references: EMBL:X63465; NID:g32079; PIDN:CAA45067.1; PID:g32080
 C:Genetics:
 A:Gene: hGDS

Query Match 60.0%; Score 39; DB 2; Length 558;
 Best Local Similarity 54.5%; Pred. No. 84;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 VFDGGDGTYYO 12
 :|||:|
 DB 260 LFEKGKGSVFO 270

RESULT 12
 JEO358

ralb binding protein - clawed frog
 C:Species: Xenopus sp. (clawed frog)
 C>Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 07-May-1999
 C:Accession: JEO358
 R:Iouzalet, N.; Camonis, J.; Moreau, J.
 Biochem. Biophys. Res. Commun. 250, 359-363, 1998
 A>Title: Identification and characterization in Xenopus of XsmgGDS, a RalB binding pr

A:Reference number: JE0358; MUID:98440802; PMID:9753634
 A:Accession: JE0358
 A:Molecule type: mRNA
 A:Residues: 1-615 <IOU>
 A:Cross-references: GB:AJ005870

Query Match 60.0%; Score 39; DB 2; Length 615;
 Best Local Similarity 54.5%; Pred. No. 93;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 OY 2 VFDDGDTGYVQ 12
 :|||1:|:
 Db 317 LFEKGKSVFQ 327

RESULT 13
 AB4265
 hypothetical protein Vng1087c [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: AB4265
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jadhav, K.H.; Alam, M.; Freitas, T.; 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li, A.; Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: AB4160; MUID:20504483; PMID:11016950
 A:Accession: AB4265
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-647 <STO>
 A:Cross-references: GB:AE004437; MUID:q10580635; PIDN:AG19485.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: VNG1087C

Query Match 59.2%; Score 38.5; DB 2; Length 647;
 Best Local Similarity 50.0%; Pred. No. 1.2e+02;
 Matches 9; Conservative 1; Mismatches 1; Indels 7; Gaps 1;
 OY 2 VFDDGDTGYVQ 12
 :|||1:|:
 Db 53 VFDDGMDRVADADGTYE 70

RESULT 14
 T40524
 hypothetical protein SPBC530.08 - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
 C:Accession: T40524
 R:Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M. submitted to the EMBL Data Library, May 1998
 A:Reference number: Z21934
 A:Accession: T40524
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-815 <LYN>
 A:Cross-references: EMBL:AL023634; PIDN:CA19174.1; GSPDB:GN00067; SPDB:SPBC530.08
 C:Genetics:
 A:Gene: SPDB:SPBC530.08
 A:Map position: 2
 C:Superfamily: GAL4 zinc binuclear cluster homology
 F:26-62/Domain: GAL4 zinc binuclear cluster homology <GL4>

Query Match 59.2%; Score 38.5; DB 2; Length 815;
 Best Local Similarity 69.2%; Pred. No. 1.5e+02;
 Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;
 OY 1 SVFDGDTGYV-YQ 12
 :|||1:|:
 Db 680 SVFDGNTGTIPYQ 692

RESULT 15
 C86187
 YVP8H12.12 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: C86187
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, Jensen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Matli, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talloker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: AB6141; MUID:21016719; PMID:11130712
 A:Accession: C86187
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-146 <STO>
 A:Cross-references: GB:AE005172; MUID:q2388570; PIDN:AA871451.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 58.5%; Score 38; DB 2; Length 146;
 Best Local Similarity 87.5%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 SVFDGDTG 8
 :|||1:|:
 Db 61 SVFDGDTG 68

Search completed: June 16, 2003, 12:10:42
 Job time : 3.47541 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:07:33 ; Search time 1.84426 seconds
(without alignments)
695.665 Million cells updates/sec

Title: US-09-986-682B-2
Perfect score: 65
Sequence: 1 SVFDCGDTVTYQ 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppaa/PCT07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	100.0	12	10	US-09-986-682B-2
2	65	100.0	455	10	US-09-986-682B-3
3	38	58.5	295	9	US-10-223-085-370
4	38	58.5	295	9	US-10-223-084-370
5	38	58.5	295	9	US-10-223-088-370
6	38	58.5	295	9	US-10-223-090-370
7	37	56.9	408	10	US-09-912-020-255
8	37	56.9	811	10	US-09-815-242-11875
9	37	56.9	1548	9	US-10-180-903-2
10	37	56.9	2076	10	US-09-815-242-5815
11	37	56.9	2186	10	US-09-815-242-12913
12	36	55.4	37	9	US-09-981-876-230
13	36	55.4	37	9	US-09-148-545-230
14	36	55.4	113	10	US-09-840-459-68
15	36	55.4	375	10	US-09-815-242-13314
16	35	53.8	4303	9	US-09-904-968A-2
17	35	53.8	140	9	US-10-101-464A-688
18	35	53.8	361	9	US-09-738-626-3735
19	35	53.8	378	9	US-10-260-877-48

20	35	53.8	382	10	US-09-815-242-10502	Sequence 10502, A
21	35	53.8	426	10	US-09-815-242-10569	Sequence 10569, A
22	35	53.8	499	9	US-09-738-626-4342	Sequence 4342, Ap
23	35	53.8	503	10	US-09-990-337-3	Sequence 3, Appl
24	35	53.8	589	12	US-10-001-851-26	Sequence 26, Appl
25	35	53.8	618	12	US-10-005-647-2	Sequence 2, Appl
26	35	53.8	627	9	US-10-050-704-125	Sequence 125, App
27	35	53.8	953	9	US-09-884-696-3	Sequence 3, Appl
28	35	53.8	956	9	US-09-884-696-4	Sequence 4, Appl
29	35	53.8	998	9	US-10-101-664A-895	Sequence 895, App
30	35	53.8	1028	9	US-09-947-063-13	Sequence 13, Appl
31	35	53.8	1028	9	US-09-947-063-14	Sequence 14, Appl
32	35	53.8	1311	9	US-10-103-377C-6	Sequence 6, Appl
33	35	53.8	4349	9	US-10-160-758-15	Sequence 15, Appl
34	34	52.3	79	10	US-09-939-980A-403	Sequence 403, App
35	34	52.3	81	10	US-09-815-242-12719	Sequence 12719, A
36	34	52.3	133	12	US-10-006-773-9	Sequence 9, Appl
37	34	52.3	255	9	US-09-898-837A-36	Sequence 36, Appl
38	34	52.3	260	10	US-09-731-231A-5	Sequence 5, Appl
39	34	52.3	267	10	US-09-815-242-5773	Sequence 5773, Ap
40	34	52.3	283	10	US-09-952-689-5	Sequence 5393, Ap
41	34	52.3	303	10	US-09-815-242-5393	Sequence 12300, A
42	34	52.3	306	10	US-09-815-242-12300	Sequence 230, App
43	34	52.3	311	9	US-09-988-626-230	Sequence 230, App
44	34	52.3	311	9	US-09-988-687-230	Sequence 20, Appl
45	34	52.3	311	9	US-10-190-279-20	

ALIGNMENTS

RESULT 1
US-09-986-682B-2
Sequence 2, Appl
Patent No. US20020115182A1
GENERAL INFORMATION:
APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU
KINKYUJO
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/986,682B
FILING DATE: 06-NO. US20020115182A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
TOPOLOGY: linear
FRAGMENT TYPE: internal fragment

SEQUENCE DESCRIPTION: SEQ ID NO: 2;
US-09-986-682B-2
Query Match 100.0%; Score 65; DB 10; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SVFDGDDGTIVYQ 12
Db 1 SVFDGDDGTIVYQ 12
RESULT 2
US-09-986-682B-3
Sequence 3, Application US/09986682B
Patent No. US20020115182A1
GENERAL INFORMATION:
APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU
KENRYUO
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/986, 682B
FILING DATE: 06-NO. US20020115182A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI=2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 3;
US-09-986-682B-3
Query Match 100.0%; Score 65; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SVFDGDDGTIVYQ 12
Db 201 SVFDGDDGTIVYQ 212
RESULT 3
US-10-223-085-370
Sequence 370, Application US/102233085
Publication No. US20030100497A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Ferrara, Napoleone

APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Marsters, Scot A.
APPLICANT: Pan, James
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Williams, P. Mickey
APPLICANT: Ye, Weilian
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: TREATMENT OF DISORDERS INVOLVING ANGIOGENESIS
FILE REFERENCE: P3235P1C10
CURRENT APPLICATION NUMBER: US/10/223,085
CURRENT FILING DATE: 2002-08-16
PRIOR APPLICATION NUMBER: US 10/081,056
PRIOR FILING DATE: 2002-02-20
PRIOR APPLICATION NUMBER: US 60/213,637
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/219,556
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: US 60/220,624
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/220,664
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: US 60/222,695
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: US 09/643,657
PRIOR FILING DATE: 2000-08-17
PRIOR APPLICATION NUMBER: PCT/US00/23522
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US00/23328
PRIOR FILING DATE: 2000-08-24
Remaining Prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 383
SEQ ID NO 370
LENGTH: 295.
TYPE: PRT
ORGANISM: Homo sapiens
US-10-223-085-370
Query Match 58.5%; Score 38; DB 9; Length 295;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 3 FDGDDGTIV 10
Db 58 FDGDDGTIV 65
RESULT 4
US-10-223-084-370
Sequence 370, Application US/102233084
Publication No. US20030105011A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Ferrara, Napoleone
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Marsters, Scot A.
APPLICANT: Pan, James
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.

```
APPLICANT: Williams, P. Mickey
APPLICANT: Ye, Weilian
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: TREATMENT OF DISORDERS INVOLVING ANGIOGENESIS
FILE REFERENCE: P3235PIC5
CURRENT APPLICATION NUMBER: US/10/223,084
CURRENT FILING DATE: 2002-08-16
PRIOR APPLICATION NUMBER: US 10/081,056
PRIOR FILING DATE: 2002-02-20
PRIOR APPLICATION NUMBER: US 60/213,637
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/219,556
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: US 60/220,624
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/220,664
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: US 60/222,695
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: US 09/643,657
PRIOR FILING DATE: 2000-08-17
PRIOR APPLICATION NUMBER: PCT/US00/23522
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US00/23328
PRIOR FILING DATE: 2000-08-24
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 383
SEQ ID NO 370
LENGTH: 295
TYPE: PRT
ORGANISM: Homo sapiens
US-10-223-084-370
```

```
Query Match      58.5%; Score 38; DB %; Length 295;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      3 FDGSDGTV 10
        ||||| 1:
Db      58 FDGSDATI 65
```

RESULT 5

```
US-10-223-088-370
Sequence 370, Application US/10223088
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Ferrara, Napoleone
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Marsters, Scot A.
APPLICANT: Pan, James
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Williams, P. Mickey
APPLICANT: Ye, Weilian
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: TREATMENT OF DISORDERS INVOLVING ANGIOGENESIS
FILE REFERENCE: P3235PIC6
CURRENT APPLICATION NUMBER: US/10/223,088
CURRENT FILING DATE: 2002-08-16
PRIOR APPLICATION NUMBER: US 10/081,056
PRIOR FILING DATE: 2002-02-20
PRIOR APPLICATION NUMBER: US 60/213,637
PRIOR FILING DATE: 2000-06-23
```

```
PRIOR APPLICATION NUMBER: US 60/219,556
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: US 60/220,624
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/220,664
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: US 60/222,695
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: US 09/643,657
PRIOR FILING DATE: 2000-08-17
PRIOR APPLICATION NUMBER: PCT/US00/23522
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US00/23328
PRIOR FILING DATE: 2000-08-24
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 383
SEQ ID NO 370
LENGTH: 295
TYPE: PRT
ORGANISM: Homo sapiens
US-10-223-088-370
```

```
Query Match      58.5%; Score 38; DB %; Length 295;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      3 FDGSDGTV 10
        ||||| 1:
Db      58 FDGSDATI 65
```

RESULT 6

```
US-10-223-090-370
Sequence 370, Application US/10223090
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Ferrara, Napoleone
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Marsters, Scot A.
APPLICANT: Pan, James
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Williams, P. Mickey
APPLICANT: Ye, Weilian
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: TREATMENT OF DISORDERS INVOLVING ANGIOGENESIS
FILE REFERENCE: P3235PIC2
CURRENT APPLICATION NUMBER: US/10/223,090
CURRENT FILING DATE: 2002-08-16
PRIOR APPLICATION NUMBER: US 10/081,056
PRIOR FILING DATE: 2002-02-20
PRIOR APPLICATION NUMBER: US 60/213,637
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/219,556
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: US 60/220,624
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: US 60/222,695
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: US 09/643,657
```

;; PRIOR FILING DATE: 2000-08-17
;; PRIOR APPLICATION NUMBER: PCT/US00/23522
;; PRIOR FILING DATE: 2000-08-23
;; PRIOR APPLICATION NUMBER: PCT/US00/23328
;; PRIOR FILING DATE: 2000-08-24
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 383
;; SEQ ID NO 370
;; LENGTH: 295
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-223-090-370

Query Match 58.5%; Score 38; DB 9; Length 295;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 FDGDDGTIV 10
Db 58 FDGDDATI 65

RESULT 7
US-09-912-020-255
;; Sequence 255, Application US/09912020
;; Patent No. US20020045592A1
;; GENERAL INFORMATION:
;; APPLICANT: Ohlsen, Karl L.
;; APPLICANT: Trawick, John
;; APPLICANT: Forsyth, R. Allyn
;; APPLICANT: Froelich, Jamie M.
;; APPLICANT: Carr, Grant J.
;; APPLICANT: Yamamoto, Robert T.
;; APPLICANT: Xu, H. Howard
;; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;; FILE REFERENCE: ELITRA.001DV1
;; CURRENT APPLICATION NUMBER: US/09/912,020
;; PRIOR FILING DATE: 2001-07-23
;; PRIOR APPLICATION NUMBER: 09/492,709
;; PRIOR FILING DATE: 2000-01-27
;; PRIOR APPLICATION NUMBER: 60/117,405
;; PRIOR FILING DATE: 1999-01-27
;; NUMBER OF SEQ ID NOS: 485
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 255
;; LENGTH: 408
;; TYPE: PRT
;; ORGANISM: E. Coli
US-09-912-020-255

Query Match 56.9%; Score 37; DB 10; Length 408;
Best Local Similarity 77.8%; Pred. No. 2.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VPDGDDGTIV 10
Db 68 VADGGEGTV 76

RESULT 8
US-09-815-242-11875
;; Sequence 11875, Application US/09815242
;; Patent No. US20020061569A1
;; GENERAL INFORMATION:
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Karl L.
;; APPLICANT: Zyskind, Judith W.
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John D.
;; APPLICANT: Carr, Grant J.
;; APPLICANT: Yamamoto, Robert T.

;; APPLICANT: Xu, H. Howard
;; TITLE OF INVENTION: Identification of Essential Genes in
;; FILE REFERENCE: ELITRA.011A
;; CURRENT APPLICATION NUMBER: US/09/815,242
;; CURRENT FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 1410
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 11875
;; LENGTH: 811
;; TYPE: PRT
;; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11875

Query Match 56.9%; Score 37; DB 10; Length 811;
Best Local Similarity 85.7%; Pred. No. 5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 FDGDDGT 9
Db 734 FDGDDGS 740

RESULT 9
US-10-180-903-2
;; Sequence 2, Application US/10180903
;; Publication No. US2003009382A1
;; GENERAL INFORMATION:
;; APPLICANT: Allen, Keith D.
;; TITLE OF INVENTION: SPC6 SERINE PROTEASE GENE DISRUPTIONS,
;; FILE REFERENCE: R-720
;; CURRENT APPLICATION NUMBER: US/10/180,903
;; CURRENT FILING DATE: 2002-06-25
;; PRIOR APPLICATION NUMBER: US 60/300,978
;; PRIOR FILING DATE: 2001-06-26
;; PRIOR APPLICATION NUMBER: US 60/324,820
;; PRIOR FILING DATE: 2001-09-24
;; NUMBER OF SEQ ID NOS: 4
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 2
;; LENGTH: 1548
;; TYPE: PRT
;; ORGANISM: Mus Musculus
US-10-180-903-2

Query Match 56.9%; Score 37; DB 9; Length 1548;
Best Local Similarity 63.6%; Pred. No. 9.9e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 VPDGDDGTIVQ 12
Db 1513 VYWGDDGTIVR 1523

RESULT 10
US-09-815-242-5815
;; Sequence 5815, Application US/09815242
;; Patent No. US20020061569A1

```

: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 5815
: LENGTH: 2076
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
US-09-815-242-5815

Query Match          56.9%; Score 37; DB 10; Length 2076;
Best Local Similarity 58.3%; Pred. No. 1.3e+03;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      1  SVFDGGDGTVYQ 12
       1:1111111111
Db      291  SMFDGDKNKYQ 302

RESULT 11
US-09-815-242-12913
: Sequence 12913, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 5815
: LENGTH: 2076
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
US-09-815-242-5815
```

```

: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 12913
: LENGTH: 2186
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
US-09-815-242-12913

Query Match          56.9%; Score 37; DB 10; Length 2186;
Best Local Similarity 58.3%; Pred. No. 1.4e+03;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      1  SVFDGGDGTVYQ 12
       1:1111111111
Db      291  SMFDGDKNKYQ 302

RESULT 12
US-09-981-876-230
: Sequence 230, Application US/09981876
: Patent No. US20020164669A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: 70 Human Secreted Proteins
: FILE REFERENCE: PZ001P1
: CURRENT APPLICATION NUMBER: US/09/981,876
: PRIOR FILING DATE: 2001-10-19
: PRIOR APPLICATION NUMBER: 09/148,545
: PRIOR FILING DATE: 1998-09-04
: PRIOR APPLICATION NUMBER: 60/040,162
: PRIOR FILING DATE: 1997-03-07
: PRIOR APPLICATION NUMBER: 60/040,333
: PRIOR FILING DATE: 1997-03-07
: PRIOR APPLICATION NUMBER: 60/038,621
: PRIOR FILING DATE: 1997-03-07
: PRIOR APPLICATION NUMBER: 60/040,161
: PRIOR FILING DATE: 1997-03-07
: PRIOR APPLICATION NUMBER: 60/040,626
: PRIOR FILING DATE: 1997-03-07
: PRIOR APPLICATION NUMBER: 60/040,334
: PRIOR FILING DATE: 1997-03-07
: PRIOR APPLICATION NUMBER: 60/040,336
: PRIOR FILING DATE: 1997-03-07
: PRIOR APPLICATION NUMBER: 60/040,163
: PRIOR FILING DATE: 1997-03-07
: PRIOR APPLICATION NUMBER: 60/047,615
: PRIOR FILING DATE: 1997-05-23
: PRIOR APPLICATION NUMBER: 60/047,600
: PRIOR FILING DATE: 1997-05-23
: PRIOR APPLICATION NUMBER: 60/047,597
: PRIOR FILING DATE: 1997-05-23
: PRIOR APPLICATION NUMBER: 60/047,502
: PRIOR FILING DATE: 1997-05-23
: PRIOR APPLICATION NUMBER: 60/047,633
: PRIOR FILING DATE: 1997-05-23
: PRIOR APPLICATION NUMBER: 60/047,583
: PRIOR FILING DATE: 1997-05-23
: PRIOR APPLICATION NUMBER: 60/047,617
: PRIOR FILING DATE: 1997-05-23
: PRIOR APPLICATION NUMBER: 60/047,618
: PRIOR FILING DATE: 1997-05-23
: PRIOR APPLICATION NUMBER: 60/047,503
: PRIOR FILING DATE: 1997-05-23
: PRIOR APPLICATION NUMBER: 60/047,592
: PRIOR FILING DATE: 1997-05-23
: PRIOR APPLICATION NUMBER: 60/047,581
: PRIOR FILING DATE: 1997-05-23
: PRIOR APPLICATION NUMBER: 60/047,584
: PRIOR FILING DATE: 1997-05-23
```

PRIOR APPLICATION NUMBER: 60/047, 500	PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: 60/047, 567	PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: 60/047, 492	PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: 60/047, 598	PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: 60/047, 613	PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: 60/047, 582	PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: 60/047, 596	PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: 60/047, 612	PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: 60/047, 632	PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: 60/047, 601	PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: 60/043, 580	PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: 60/043, 568	PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: 60/043, 674	PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: 60/043, 569	PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: 60/043, 311	PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: 60/043, 671	PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: 60/043, 674	PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: 60/043, 672	PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: 60/043, 315	PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: 60/048, 974	PRIOR FILING DATE: 1997-06-06
PRIOR APPLICATION NUMBER: 60/056, 886	PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056, 877	PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056, 889	PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056, 893	PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056, 630	PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056, 878	PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056, 662	PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056, 872	PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056, 882	PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056, 637	PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056, 903	PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056, 888	PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056, 879	PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056, 880	PRIOR FILING DATE: 1997-08-22

1	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 8894
2	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 9111
3	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 9536
4	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 8784
5	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 9100
6	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 8684
7	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 6311
8	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 8451
9	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 8932
10	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/047, 5595
11	PRIOR FILING DATE:	1997-05-23	PRIOR APPLICATION NUMBER:	60/057, 7611
12	PRIOR FILING DATE:	05-Sep-1997	PRIOR APPLICATION NUMBER:	60/047, 5599
13	PRIOR FILING DATE:	1997-05-23	PRIOR APPLICATION NUMBER:	60/047, 5588
14	PRIOR FILING DATE:	1997-05-23	PRIOR APPLICATION NUMBER:	60/047, 5855
15	PRIOR FILING DATE:	1997-05-23	PRIOR APPLICATION NUMBER:	60/047, 5866
16	PRIOR FILING DATE:	1997-05-23	PRIOR APPLICATION NUMBER:	60/047, 5900
17	PRIOR FILING DATE:	1997-05-23	PRIOR APPLICATION NUMBER:	60/047, 5944
18	PRIOR FILING DATE:	1997-05-23	PRIOR APPLICATION NUMBER:	60/047, 5699
19	PRIOR FILING DATE:	1997-05-23	PRIOR APPLICATION NUMBER:	60/047, 5933
20	PRIOR FILING DATE:	1997-05-23	PRIOR APPLICATION NUMBER:	60/047, 5766
21	PRIOR FILING DATE:	1997-05-23	PRIOR APPLICATION NUMBER:	60/047, 6144
22	PRIOR FILING DATE:	1997-05-23	PRIOR APPLICATION NUMBER:	60/043, 5788
23	PRIOR FILING DATE:	1997-04-11	PRIOR APPLICATION NUMBER:	60/043, 5766
24	PRIOR FILING DATE:	1997-04-11	PRIOR APPLICATION NUMBER:	60/047, 5010
25	PRIOR FILING DATE:	1997-05-23	PRIOR APPLICATION NUMBER:	60/043, 6700
26	PRIOR FILING DATE:	1997-04-11	PRIOR APPLICATION NUMBER:	60/056, 6322
27	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 6654
28	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 8766
29	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 8811
30	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 9090
31	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 8755
32	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 8682
33	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 8887
34	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 9080
35	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/048, 9654
36	PRIOR FILING DATE:	1997-06-06	PRIOR APPLICATION NUMBER:	60/057, 6500
37	PRIOR FILING DATE:	1997-09-05	PRIOR APPLICATION NUMBER:	60/056, 8884
38	PRIOR FILING DATE:	1997-08-22	PRIOR APPLICATION NUMBER:	60/056, 8894

Query Match: 55.48; Score 36; DB 9; Length 37;
Best Local Similarity 75.08; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 4 DGGDCTVY 11
|||||
Db 25 DGGDCTSF 32

RESULT 13
US-09-148-545-230
Sequence 230, Application US/09148545
Publication No. US20030027132A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 70 Human Secreted Proteins
FILE REFERENCE: P2001P1
CURRENT APPLICATION NUMBER: US/09/148,545
EARLIER FILING DATE: 1998-09-04
EARLIER APPLICATION NUMBER: PCT/US98/04482
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,161
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598

EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,315
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/056,886
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,879
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,880
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,894
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,911
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,636
EARLIER FILING DATE: 1997-08-22

EARLIER APPLICATION NUMBER: 60/056,874
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,910
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,864
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,631
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,845
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,892
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/047,595
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/057,761
EARLIER FILING DATE: 05-Sep-1997
EARLIER APPLICATION NUMBER: 60/047,599
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,588
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,585
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,586
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,590
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,594
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,589
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,593
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,614
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,578
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,670
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056,632
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,664
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
NUMBER OF SEQ ID NOS: 280
SOFTWARE: Patenlin Ver. 2.0
SEQ ID NO 230
LENGTH: 37
Query Match 55.4%; Score 36; DB 9; Length 37;
Best Local Similarity 75.0%; Pred. No. 28;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 4 DGGDGTGY 11
|||||
Db 25 DGGDGTSF 32
RESULT 14
US-09-840-459-68
; Sequence 68, Application US/09840459
; Patent No. US20020150576A1
; GENERAL INFORMATION:
; APPLICANT: Lakosa, Gregory J.
; APPLICANT: Horvath, Christopher
; APPLICANT: Newman, Walter
; APPLICANT: Jones, S. Tarran
; APPLICANT: O'Brien, Theresa
; TITLE OF INVENTION: HUMANIZED ANTI-CCR2 ANTIBODIES AND
; FILE REFERENCE: 1855,1052-012
; CURRENT APPLICATION NUMBER: US/09/840,459
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: PCT/US01/03537
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/497,625
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 09/359,193
; PRIOR FILING DATE: 1999-07-22
; PRIOR APPLICATION NUMBER: 09/121,781
; PRIOR FILING DATE: 1998-07-23
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 68
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-840-459-68
Query Match 55.4%; Score 36; DB 10; Length 113;
Best Local Similarity 54.5%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 SVFDGDCGTGY 11
|:| ||| |
Db 28 SLDSGDCNTY 38
RESULT 15
US-09-815-242-13314
; Sequence 13314, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 13314
; LENGTH: 375
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13314

Query Match 55.4%; Score 36; DB 10; Length 375;
Best Local Similarity 77.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VPDGGDGTV 10
| | | | |
Db 45 VPDGGDGTV 53

Search completed: June 16, 2003, 12:22:36
Job time : 2.84426 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:04:38 ; Search time 1.08197 Seconds
(Without alignments)
326.327 Million cell updates/sec

Title: US-09-986-682B-2
Perfect score: 65
Sequence: 1 SVFDGSDGTIVYQ 12

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	100.0	12	US-08-870-827-2	Sequence 2, Appli
2	65	100.0	12	US-09-317-179-2	Sequence 2, Appli
3	65	100.0	455	US-08-870-827-3	Sequence 3, Appli
4	65	100.0	455	US-09-317-179-3	Sequence 3, Appli
5	37	56.9	190	US-08-560-398-8	Sequence 8, Appli
6	37	56.9	1018	US-08-408-093-6	Sequence 6, Appli
7	37	56.9	1018	US-08-408-420A-6	Sequence 6, Appli
8	37	56.9	1018	US-08-714-901-6	Sequence 6, Appli
9	37	56.9	1018	US-08-452-052-2	Sequence 2, Appli
10	37	56.9	1018	US-08-040-741-6	Sequence 6, Appli
11	36	55.4	460	US-08-289-709-1	Sequence 1, Appli
12	36	55.4	460	US-08-602-656-1	Sequence 1, Appli
13	36	55.4	4302	US-08-658-136-5	Sequence 1, Appli
14	36	55.4	4302	US-09-052-469-8	Sequence 8, Appli
15	36	55.4	4303	US-08-460-751-2	Sequence 2, Appli
16	36	55.4	4339	US-09-052-469-6	Sequence 6, Appli
17	35	53.8	140	US-08-930-894-4	Sequence 4, Appli
18	35	53.8	249	US-08-155-171B-32	Sequence 32, Appli
19	35	53.8	249	US-08-435-998-32	Sequence 32, Appli
20	35	53.8	512	US-08-173-508-4	Sequence 4, Appli
21	35	53.8	512	US-08-265-310-4	Sequence 4, Appli
22	35	53.8	512	US-08-951-742-4	Sequence 4, Appli
23	35	53.8	530	US-08-943-714-12	Sequence 12, Appli
24	35	53.8	956	US-08-772-270A-8	Sequence 8, Appli
25	35	53.8	1265	US-09-347-878-5	Sequence 5, Appli
26	35	53.8	1265	US-09-347-878-7	Sequence 7, Appli
27	34	52.3	79	US-08-936-165A-403	Sequence 403, App

28	34	52.3	93	2	US-08-341-843B-32	Sequence 32, Appli
29	34	52.3	93	2	US-08-427-497E-37	Sequence 37, Appli
30	34	52.3	238	4	US-09-605-858-35	Sequence 35, Appli
31	34	52.3	265	3	US-08-483-857-8	Sequence 8, Appli
32	34	52.3	270	4	US-09-134-001C-3942	Sequence 3942, Ap
33	34	52.3	307	4	US-09-134-001C-3238	Sequence 3238, Ap
34	34	52.3	311	4	US-09-564-805-230	Sequence 230, App
35	34	52.3	312	4	US-09-134-001C-5583	Sequence 5583, Ap
36	34	52.3	326	4	US-09-564-805-231	Sequence 231, App
37	34	52.3	448	4	US-09-627-650B-17	Sequence 17, Appli
38	34	52.3	448	4	US-09-436-063C-17	Sequence 17, Appli
39	34	52.3	475	4	US-09-627-650B-16	Sequence 16, Appli
40	34	52.3	475	4	US-09-436-063C-16	Sequence 16, Appli
41	34	52.3	487	4	US-09-627-650B-13	Sequence 13, Appli
42	34	52.3	487	4	US-09-627-650B-14	Sequence 14, Appli
43	34	52.3	487	4	US-09-436-063C-13	Sequence 13, Appli
44	34	52.3	487	4	US-09-436-063C-14	Sequence 14, Appli
45	34	52.3	506	4	US-09-627-650B-15	Sequence 15, Appli

ALIGNMENTS

```
RESULT 1
US-08-870-827-2
: Sequence 2, Application US/08870827
: Patent No. 5962297
:
: GENERAL INFORMATION:
: APPLICANT: Tsusaki et al.
: TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
: TITLE OF INVENTION: ACTIVITY
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/870,827
: FILING DATE: 06-JUN-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 170,630/1996
: FILING DATE: 10-JUN-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: TSUSAKI-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 12 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
:
: FRAGMENT TYPE: internal fragment
:
: US-08-870-827-2
:
: Query Match 100.0%; Score 65; DB 2; Length 12;
: Best Local Similarity 100.0%; Pred. No. 8e-05;
: Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: QY 1 SVFDGSDGTIVYQ 12
: |||||||
: Db 1 SVFDGSDGTIVYQ 12
```

RESULT 2
US-09-317-179-2
Sequence 2, Application US/09317179
Patent No. 6383769
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/317,179
FILING DATE: 24-May-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/870,827
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
TOPOLOGY: linear
FRAGMENT TYPE: Internal fragment
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-317-179-2
Query Match 100.0%; Score 65; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SVFDGSGDTVYQ 12
Db 1 SVFDGSGDTVYQ 12
RESULT 3
US-08-870-827-3
Sequence 3, Application US/08870827
Patent No. 5962297
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,827
FILING DATE: 06-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-870-827-3
Query Match 100.0%; Score 65; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SVFDGSGDTVYQ 12
Db 201 SVFDGSGDTVYQ 212
RESULT 4
US-09-317-179-3
Sequence 3, Application US/09317179
Patent No. 6383769
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/317,179
FILING DATE: 24-May-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/870,827
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-317-179-3

Query Match 100.0%; Score 65; DB 4; Length 455;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SVFDGDTYVQ 12
|||||
DB 201 SVFDGDTYVQ 212

RESULT 5
US-08-560-398-8
Sequence 8, Application US/08560398
Patent No. 5807082

GENERAL INFORMATION:
APPLICANT: O'Neill, Sharmen
APPLICANT: Nadeau, Jeanette
TITLE OF INVENTION: Ovale-Specific Gene Expression
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/560.398
FILING DATE: 17-NOV-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-06330005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 190 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-560-398-8

Query Match 56.9%; Score 37; DB 2; Length 190;
Best Local Similarity 85.7%; Pred. No. 54;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 FDGSDGT 9
|||||
DB 68 FDGSDGS 74

RESULT 6
US-08-408-093-6
Sequence 6, Application US/08408093
Patent No. 5688916

GENERAL INFORMATION:

APPLICANT: Reid, Robert A.
APPLICANT: Hemperly, John J.
TITLE OF INVENTION: Human Cell Adhesion Molecule and Nucleic
Acid Sequences
NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Richard J. Rodrick, Becton Dickinson
ADDRESSEE: and Company
STREET: One Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: USA
ZIP: 07417

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/408.093
FILING DATE: 21-MAR-1995
CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/040.741
FILING DATE: 26 MAR 1993
ATTORNEY/AGENT INFORMATION:

NAME: Pugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-2630

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1018 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

ORIGINAL SOURCE:
ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 45..94

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 138..191

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 243..290

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 332..371

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 416..464

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 506..563

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 604..657

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 707..760

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 809..857

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 905..952

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 953..1000

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 1001..1050

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 1051..1100

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 1101..1150

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 1151..1200

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 1201..1250

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 1251..1300

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 1301..1350

OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 188
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 238
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 318
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 437
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 453
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 474
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 501
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 571
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 913
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
US-08-408-093-6

Query Match 56.9%; Score 37; DB 1; length 1018;
Best Local Similarity 77.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 DGGDGVYQ 12
| | | | |
Db 981 DGGDGVYQ 989

RESULT 7
US-08-408-420A-6
Sequence 6, Application US/08408420A
Patent No. 5731154
GENERAL INFORMATION:
APPLICANT: Reid, Robert A.
APPLICANT: Hemperly, John J.

TITLE OF INVENTION: Human Cell Adhesion Molecule and Nucleic
TITLE OF INVENTION: Acid Sequences
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
STREET: One Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: US
ZIP: 07417
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/408,420A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-2630
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1018 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 45..94
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 138..191
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 243..290
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 332..371
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 416..464
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 506..563
FEATURE:
NAME/KEY: Domain
LOCATION: 604..657
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note="conserved core of fibronectin type
OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 707..760
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note="conserved core of fibronectin type
OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 809..857
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note="conserved core of fibronectin type
OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 905..952
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note="conserved core of fibronectin type

OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 188
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 238
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 318
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 437
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 453
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 474
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 501
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 571
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 913
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
US-08-408-420A-6

Query Match 56.9%; Score 37; DB 1; Length 1018;
Best Local Similarity 77.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 DGGDGTYYQ 12
DB 981 DGGDGTYYQ 989

RESULT 8
US-08-714-901-6
Sequence 6, Application US/08714901
Patent No. 5739289
GENERAL INFORMATION:
APPLICANT: Reid, Robert A.
APPLICANT: Hemperly, John J.

TITLE OF INVENTION: Human Cell Adhesion Molecule and Nucleic
TITLE OF INVENTION: Acid Sequences
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson
ADDRESSEE: and Company
STREET: One Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: USA
ZIP: 07417
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,901
FILING DATE: 17-SEP-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/408,093
FILING DATE: 21-MAR-1995
APPLICATION NUMBER: US/08/040,741
FILING DATE: 26 MAR 1993
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-2630
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1018 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo. sapiens
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 45..94
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 138..191
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 243..290
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 332..371
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 416..464
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 506..563
FEATURE:
NAME/KEY: Domain
LOCATION: 604..657
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 707..760
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 809..857
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIT-like repeat"

FEATURE:
NAME/KEY: Domain
LOCATION: 905..952
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: tit-like repeat"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 188
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 238
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 318
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 437
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 453
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 474
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 501
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 571
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 913
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
US-08-714-901-6

Query Match 56.9%; Score 37; DB 1; Length 1018;
Best Local Similarity 77.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 DGGDGTVO 12
|||||
DB 981 DGGDGVVSO 989

RESULT 9
US-08-452-052-2

; Sequence 2, Application US/08452052
; Patent No. 5766922
; GENERAL INFORMATION:
; APPLICANT: Peles, Elior
; TITLE OF INVENTION: FUNCTIONAL LIGANDS FOR THE AXONAL CELL
; TITLE OF INVENTION: RECOGNITION MOLECULE CONTACTIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,052
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7683-075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-452-052-2

Query Match 56.9%; Score 37; DB 1; Length 1018;
Best Local Similarity 77.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 DGGDGTVO 12
|||||
DB 980 DGGDGVVSO 988

RESULT 10
US-08-040-741-6
; Sequence 6, Application US/08040741
; Patent No. 6017695
; GENERAL INFORMATION:
; APPLICANT: Reid, Robert A.
; APPLICANT: Hemperly, John J.
; TITLE OF INVENTION: Human Cell Adhesion Molecule and Nucleic
; TITLE OF INVENTION: Acid Sequences
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson
; ADDRESSEE: and Company
; STREET: One Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: USA
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/040,741
FILING DATE: 19930326
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-2630
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1018 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 45..94
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 138..191
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 243..290
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 332..371
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 416..464
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 506..563
FEATURE:
NAME/KEY: Domain
LOCATION: 604..657
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIT-1like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 707..760
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIT-1like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 809..857
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIT-1like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 905..952
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIT-1like repeat"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 188
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 238
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 318
OTHER INFORMATION: /label= ASN-glycos

OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 437
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 453
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 474
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 501
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 571
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 913
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
OTHER INFORMATION: glycosylation"
US-08-040-741-6
Query Match 56.9%; Score 37; DB 3; Length 1018;
Best Local Similarity 77.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Caps 0;
QY 4 DGGDGYQYQ 12
DB 981 DGGDGVVQ 989
RESULT 11
US-08-289-709-1
Sequence 1, Application US/08289709
Patent No. 5523224
GENERAL INFORMATION:
APPLICANT: Burtischer, Helmut; Lang, Gunter; Popp, Friedrich
TITLE OF INVENTION: Recombinant D-Hydantoinase, A Process for the
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,709
FILING DATE: 12-AUGUST-1994

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P 43 28 829.4
FILING DATE: 27-AUGUST-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5523224man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: BOER 1041
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-289-709-1

Query Match 55.4%; Score 36; DB 1; Length 460;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 SVFDGGDTVQ 12
DB 157 NVFOADDGTVR 168

RESULT 12
US-08-602-656-1
Sequence 1, Application US/08602656
Patent No. 5679571
GENERAL INFORMATION:
APPLICANT: Bartscher, Helmut; Lang, Gunter; Popp, Friedrich
TITLE OF INVENTION: Recombinant D-Hydantoinase, A Process For the
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,656
FILING DATE: 16-FEB-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/289,709
FILING DATE: 12-AUGUST-1994
APPLICATION NUMBER: P 43 28 829.4
FILING DATE: 27-AUGUST-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5679571man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: BOER 1041
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-602-656-1

Query Match 55.4%; Score 36; DB 1; Length 460;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 SVFDGGDTVQ 12
DB 157 NVFOADDGTVR 168

RESULT 13
US-08-658-136-5
Sequence 5, Application US/08658136
Patent No. 6071717
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W
APPLICANT: LANDES, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY
APPLICANT: OLAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN4-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-5415
TELEFAX: 508-872-5415
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-658-136-5

Query Match 55.4%; Score 36; DB 3; Length 4302;
Best Local Similarity 87.5%; Pred. No. 2.1e+03;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 FDGGDTV 10
DB 1245 FDMGDTV 1252

RESULT 14
US-09-052-469-8
Sequence 8, Application US/09052469
Patent No. 6380360
GENERAL INFORMATION:
APPLICANT: Harris et al.
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Wilcoff, Ltd.
STREET: One Financial Center
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk, 3.50 Inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/052,469
FILING DATE: Concurrently herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/422,582
FILING DATE: 14-April-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9507766.5
FILING DATE: 13-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9411900.5
FILING DATE: 14-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB PCT/GB94/02822
FILING DATE: 23-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9326470.3
FILING DATE: 24-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Ph.D., Kathleen M.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: 3265/74165
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 345-9100
TELEFAX: (617) 345-9111
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-052-469-8
Query Match 55.4%; Score 36; DB 4; Length 4302;
Best Local Similarity 87.5%; Pred. NO. 2.1e+03;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 FDGSDGTV 10
Db 1245 FDMGSDGTV 1252
RESULT 15
US-08-460-751-2
Sequence 2, Application US/08460751
Patent No. 5891628
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen
APPLICANT: Schneider, Michael
APPLICANT: Glucksmann, Sandra
TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
DISEASE GENE, DIAGNOSTICS AND TREATMENT
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,751
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/413,580
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7638-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4303 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-460-751-2
Query Match 55.4%; Score 36; DB 2; Length 4303;
Best Local Similarity 87.5%; Pred. NO. 2.1e+03;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 FDGSDGTV 10
Db 1245 FDMGSDGTV 1252

Search completed: June 16, 2003, 12:11:32
Job time : 2.08197 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:55:48 ; Search time 121.209 seconds
(Without alignments)
500.202 Million cell updates/sec

Title: US-09-986-682B-3
Perfect score: 2454
Sequence: 1 MNSGDYKEDYGFHITRADM.....NIKSDTSGSENSGQGF 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2454	100.0	487	19	AAW48306	Bacillus sp. V230
2	1553	63.3	527	22	ABG04307	Novel human diagno
3	1553	63.3	644	22	ABG06230	Novel human diagno
4	1553	63.3	736	22	ABG22589	Novel human diagno
5	1553	63.3	774	22	ABG05590	Novel human diagno
6	1553	63.3	774	22	ABG11754	Novel human diagno
7	1553	63.3	789	22	ABG04946	Novel human diagno
8	1553	63.3	789	22	ABG23205	Novel human diagno
9	1553	63.3	789	22	ABG28407	Novel human diagno
10	1553	63.3	823	22	ABG25769	Novel human diagno

11	1553	63.3	855	22	ABG04500	Novel human diagno
12	1553	63.3	881	22	ABG21651	Novel human diagno
13	1553	63.3	893	22	ABG07167	Novel human diagno
14	1553	63.3	913	22	ABG02461	Novel human diagno
15	1553	63.3	929	22	ABG21199	Novel human diagno
16	1553	63.3	1027	22	ABG21645	Novel human diagno
17	1551	63.2	472	12	AA10671	B.amy101liquefacien
18	1551	63.2	493	22	ABG21647	Novel human diagno
19	1545	63.0	689	22	ABG00041	Novel human diagno
20	1543.5	62.9	546	22	ABG21832	Novel human diagno
21	1543.5	62.9	562	22	ABG21605	Novel human diagno
22	1543.5	62.9	1095	22	ABG21602	Novel human diagno
23	1543.5	62.9	1095	22	ABG21823	Novel human diagno
24	1510	61.5	542	22	ABG21613	Novel human diagno
25	1476.5	60.2	1551	22	ABG28409	Novel human diagno
26	1473	60.0	797	22	ABG18136	Novel human diagno
27	1471	59.9	786	22	ABG02257	Novel human diagno
28	1444	58.8	3048	22	ABG25791	Novel human diagno
29	1427.5	58.2	495	22	ABG04320	Novel human diagno
30	1378.5	56.2	657	22	ABG21612	Novel human diagno
31	1378.5	56.2	657	22	ABG21834	Novel human diagno
32	1375	56.0	436	22	ABG04298	Novel human diagno
33	1368	55.7	778	22	ABG13422	Novel human diagno
34	1362	55.5	426	22	ABG04288	Novel human diagno
35	1340.5	54.6	981	22	ABG21856	Novel human diagno
36	1339.5	54.6	893	22	ABG26977	Novel human diagno
37	1329	54.2	828	22	ABG24549	Novel human diagno
38	1329	54.2	928	22	ABG21821	Novel human diagno
39	1329	54.2	928	22	ABG22394	Novel human diagno
40	1329	54.2	928	22	ABG26921	Novel human diagno
41	1325.5	54.0	708	22	ABG25779	Novel human diagno
42	1325.5	54.0	708	22	ABG28401	Novel human diagno
43	1314	53.5	914	22	ABG09624	Novel human diagno
44	1309	53.3	1034	22	ABG21868	Novel human diagno
45	1307.5	53.3	665	22	ABG04325	Novel human diagno

ALIGNMENTS

RESULT 1	
AAW48306	
ID	AAW48306 standard; Protein: 487 AA.
AC	AAW48306;
XX	04-JUN-1998 (first entry)
DE	Bacillus sp. V230 beta-fructofuranosidase.
XX	
XX	Beta-fructofuranosidase; xylosyl fructoside; isomaltosyl fructoside;
KW	lactosucrose; fructosyltrehalose; anticariogenic; growth promoter;
KW	Blitidobacteria; mineral-absorption promoting activity; taste; texture;
KW	cosmetic; pharmaceutical; sweetener.
OS	Bacillus sp.
XX	
XX	
XX	Key
FT	Peptide
FT	Protein
FT	Location/Qualifiers
FT	1..32
FT	/label= signal_peptide
FT	33..487
FT	/label= beta-fructofuranosidase
FT	/note= "Claim 4"
XX	
PN	EP812915-A2.
XX	
PD	17-DEC-1997.
XX	
PF	10-JUN-1997; 97EP-0304032.
XX	
PR	10-JUN-1996; 96JP-0170630.
XX	
PA	(HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

XX Chaeen H, Kubota M, Tsusaki K:
XX
XX WPI: 1998-034976/04.
DR N-PSDB; AAV17621.
XX
XX Bacillus beta-fructofuranosidase enzyme - useful for producing
PT fructo:furanosylated saccharide(s) or alcohol(s)
PS
XX
XX Claim 1: Pages 23-25; 32pp; English.
CC This is beta-fructofuranosidase from *Bacillus* sp. V230, an enzyme that
CC has an optimum temperature of 50 deg. C at pH 6. It catalyses
CC fructofuranosyl transfer reactions from a fructofuranosyl donor to a
CC fructofuranosyl acceptor where the donor is sucrose, raffinose or erlose
CC and the acceptor is selected from alcohols, sugar alcohols and
CC saccharides having no beta-fructofuranosidic linkages,
CC especially where the reaction product is xylosyl fructoside, erlose,
CC isomaltosyl fructoside, lactosucrose or fructosyltrehalose.
CC Such reaction products have a satisfactory taste and sweetness, a
CC moderate viscosity and humectancy, an effective anticariogenic activity,
CC growth promoting activity for *Bifidobacteria*, mineral-absorption
CC promoting activity. The products can be used to improve the tastes and
CC textures of food products, cosmetics and pharmaceuticals and are useful
CC as sweeteners in the food and pharmaceutical industries.
CC
CC
SQ Sequence 487 AA:
Query Match 100.0%; Score 2454; DB 19; Length 487;
Best Local Similarity 100.0%; Pred. No. 7.6e-197;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MNSGDKYKDYGFANHTTRADMLKIPGOQNSPOFKVQFNASAKINDSAKGYDKSGNLIDL 60
DB 33 MNSGDKYKDYGFANHTTRADMLKIPGOQNSPOFKVQFNASAKINDSAKGYDKSGNLIDL 92
OY 61 DWDSWPLQNDAGTANHYGHIVSALAGDPKNSDTPLEHYQKVGDTSIDSMKAGRV 120
DB 93 DWDSWPLQNDAGTANHYGHIVSALAGDPKNSDTPLEHYQKVGDTSIDSMKAGRV 152
OY 121 FEDMDKFPNDPYLKYQTOEMSGSATLTKDGOVRLFTYDYSGNPBDGGAGNOIITSAO 180
DB 153 FEDMDKFPNDPYLKYQTOEMSGSATLTKDGOVRLFTYDYSGNPBDGGAGNOIITSAO 212
OY 181 VNLSPDAATLKVDGVSDFHKSVPFDGSGTIVYONIQOFTDEGKMWISGDHNTLDRPHYVEDK 240
DB 213 VNLSPDAATLKVDGVSDFHKSVPFDGSGTIVYONIQOFTDEGKMWISGDHNTLDRPHYVEDK 272
OY 241 GHKYLVEFANVTGTDGYOGDOSFNKKAYGSGDVFQNEKNKLLQSPKQIASLANGALG 300
DB 273 GHKYLVEFANVTGTDGYOGDOSFNKKAYGSGDVFQNEKNKLLQSPKQIASLANGALG 332
OY 301 IVELADDTYKSVKMPVLAASNTVADEVBRANIFKNNKNWYLFITDSRGSKMTSDGINDKV 360
DB 333 IVELADDTYKSVKMPVLAASNTVADEVBRANIFKNNKNWYLFITDSRGSKMTSDGINDKV 392
OY 361 YMLGCGDSLNGPNNPINETGVLVNMNLDPAADLTHTYSHCGIPHEGNNVVLTSYMTNRG 420
DB 393 YMLGCGDSLNGPNNPINETGVLVNMNLDPAADLTHTYSHCGIPHEGNNVVLTSYMTNRG 452
OY 421 FYPEHHSHLRDKLGVNIGSDTSGGESSGGCGQFP 455
DB 453 FYPEHHSHLRDKLGVNIGSDTSGGESSGGCGQFP 487

DE Novel human diagnostic protein #4298.
XX
XX Human: Chromosome mapping; gene mapping; gene therapy; forensic:
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HSE-) HSE INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI: 2001-639362/73.
DR N-PSDB; AAS68494.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20: SEQ ID NO 34666; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX
SQ Sequence 527 AA:
Query Match 63.3%; Score 1553; DB 22; Length 527;
Best Local Similarity 65.3%; Pred. No. 2.5e-121;
Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;
OY 2 NSGDKEDYGFANHTTRADMLKIPGOQNSPOFKVQFNASAKINDSAKGYDKSGNLIDL 61
DB 88 NOKAYKETYGVSHTRHMDLIPKQONEKYQVDFPDOSTKNTIESAG-----LD 138
OY 62 VMDSWPLQNDAGTANHYGHIVSALAGDPKNSDTPLEHYQKVGDTSIDSMKAGRV 121
DB 139 VMDSWPLQNDAGTANHYGHIVSALAGDPKNSDTPLEHYQKVGDTSIDSMKAGRV 198
OY 122 EDMDKFPNDPYLKYQTOEMSGSATLTKDGOVRLFTYDYSGNPBDGGAGNOIITSAO 181
DB 199 KDSKFDANDPTLKDQTOEMSGSATLTKDGOVRLFTYDYSGNPBDGGAGNOIITSAO 252
OY 182 NLSQPDAAATLKVDGVSDFHKSVPFDGSGTIVYONIQOFTDEGKMWISGDHNTLDRPHYVEDK 241
DB 253 NVSKSD-DTLKINGVEDHKITFD-GDKTYQNVQOFTDEGKMWISGDHNTLDRPHYVEDK 310

Oy		242	KKLYLFEANTGTTDVCYGDQSFNNKAYCGSVLFQFNEKNKLQSPKKQLASLANAGLG	361			
Dd		311	HKYLFEANTGTENGYGBESLEFNKAYGGTGFPRKEQOKLQOSAKRDAELANALGI	370			
Oy		302	VELADDDYVKSVMKRVLVASNTFVADVERANIFKRMNKKWYLEFTDSRSCKMTSDGINDKDY	361			
Dd		371	IELNDYTLTKVKMKRLITSNTYTDELRANNVFKMNCKWYLFITDSRSCKMTIDGINSNDIY	430			
Oy		362	MLPGEGDSLNGPHNPINETGLVANNLDPADLTHTYSHCGIPHPREGNNVLTYSVTNRGF	421			
Dd		431	MLGVYNSNLTPPKPLNKRTGLVLOGLPNDVTEFTYSHFAVPQAKGNNAVITYSTNRGF	490			
Oy		422	YREHHSHLRDLKLVNIKSGDPTSGGNS-SCGOO	453			
Dd		491	FEDKATPAFPSEFLMINKKTSVVRXNSILEGOO	523			
RESULT 3							
ID	ABG06230	standard; Protein; 644 AA.					
AC	ABG06230;						
DT	13-FEB-2002	(first entry)					
DE	Novel human diagnostic protein #6221.						
XX							
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;						
RW	food supplement; medical imaging; diagnostic; genetic disorder.						
XX							
OS	Homo sapiens.						
XX							
PN	MO200175067-A2.						
PD	11-OCT-2001.						
PE	30-MAR-2001; 2001WO-US08631.						
PR	31-MAR-2000; 2000US-0540217.						
RR	23-AUG-2000; 2000US-0649167.						
XX							
PA	(HYSE-) HYSEQ INC.						
PI	Drmanac RT, Liu C, Tang YT;						
DR	WPI: 2001-639362/73.						
XX	N-PSDB; AAS70417.						
PT	New isolated polynucleotide and encoded polypeptides, useful in						
PT	diagnostics, forensics, gene mapping, identification of mutations						
PT	responsible for genetic disorders or other traits and to assess						
XX	biodiversity -						
XX							
XS	Claim 20; SEQ ID NO 36589; 103bp; English.						
CC							
CC	The invention relates to isolated polynucleotide (I) and						
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,						
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome						
CC	and gene mapping, and in recombinant production of (II). The						
CC	polynucleotides are also used in diagnostics as expressed sequence tags						
CC	for identifying expressed genes. (I) is useful in gene therapy techniques						
CC	to restore normal activity of (II) or to treat disease states involving						
CC	(II). (II) is useful for generating antibodies against it, detecting or						
CC	quantitating a polypeptide, in tissue, as molecular weight markers and as						
CC	a food supplement. (II) and its binding partners are useful in medical						
CC	imaging of sites expressing (II). (I) and (II) are useful for treating						
CC	disorders involving aberrant protein expression or biological activity.						
CC	The polypeptide and polynucleotide sequences have applications in						
CC	diagnostics, forensics, gene mapping, identification of mutations						
CC	responsible for genetic disorders or other traits to assess biodiversity						
CC	and to produce other types of data and products dependent on DNA and						
CC	amino acid sequences. ABG00010-ABG30377 represent novel human						
CC							

CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 644 AA;

Query Match	63.3%	Score 1553	DB 22	Length 644
Best Local Similarity	65.3%	Pred. No. 3.3e-121		
Match 296: Conservative	54	Mismatches 85	Indels 18	Gaps 5

QY 2 NSGGYKEDGFGAHTITRADMLKIPGQONSQFQKVPQFNASALIKNIDSAGKGDKSGNLIDLD 61

Db 205 NQRAYKKEYGVSHITRHDMLQIPKQOQNEKYQVPEDQSTIKNIESAGK-----LD 255

QY 62 VMWSPLQNDAGTAAHYHGYHTYVSALAGDPKNSDPTPLHLFYKYGKDTSIDSWKNAAGRVF 121

Db 256 VMWSPLQNDAGTAAHYHGYHTYVSALAGDPKNSDPTPLHLFYKYGKDTSIDSWKNAAGRVF 315

QY 122 EDMKKFVAPNDPYLKYQTQEMSGSNTLLKKSQVRLEFYTDVSGNPEEGCTGAGNOILSTAQV 181
::||| ||||||||| | :||| | :|||
Db 316 KDSKSFADANDPILKDTQTEMSGSAFTSDSKIRLETYYDVS GK-----HYGQSILTTAAV 369

QY 182 NLSPDAATLLKVDGSDHKSVFEDGGDGYVONIQOFIDECKWISDNHTLRDPHYVEDKG 241
| : - : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 370 NVASDK-DLKLINGVEDHRTIFD-GDGKYIVONVOPIFIDEGYTSGSNHTLRDPHYVEDKG 427

QY 242 HKLYVEANTGTTDGYQGDOSFNKKAYGGSDVEFQNEKKLLQSPKKQTASLANCALGI 301
 |||||
 428 HKLYVEANTGTENGYYQGESLFNNAYGGGTNFEKRESOKLOQSAKKRAELANCALGI 487
 |||||

Qy 302 VELADDYATKSVKMLVATSTVADEVERANIIFKMNKKWLYFTDSGSKMTSDGINDKDY 361
 Db 488 IELNNDDYATKSVKMLVATSTVADEIERAVFKMNKKWLYFTDSGSKMTDGINSDY 547

QY 362 MLCGGDSTLNGPHNPINETGGLVLMNNLPADPLTHTYSHCIPHEGNNVLTSTYMNRGF 421
||| |||
Db 548 MLGVSVNSTLTGPYKPLNKTGLVLDMGLDPDVTFEYSHEAVPQAAGNNNVITSTYMNRGF 607

```
QY      422 YPEHSHLBDKLGVINIKGSDTSGGNS-SSQG 453
      : : : : : : : : : : : : : : : :
DB      608 FEDKKAPEAFSLMNKIKGNTSVYKNSILEQG 640
```

RESULT 4	
ABG22589	
1D	ABG22589 standard; Protein; 736 AA.
XX	
AC	ABG22589;
XX	
DT	18-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #22580.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
RW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Dzmanac RT, Liu C, Tang YT;
XX	
WP1	2001-639362/73.
XX	

Db 146 KDSKFDANDPLIKDQTOEMSGSATFTSDGKIRLFYTDYSGK -----HYGQSILTTAQV 199
QY 182 NLSQPDATLTKVDGYSVDRKSVFDDGDTYVONIQQFIDEGKWIISGDNHTLRDPHYVEDKG 241
Db 200 NVSKSD-DTLKINGVEDHKTIFD-GDGKTYQVQVOQFIDEGANTSDNHTLRDPHYVEDKG 257
QY 242 HKYLVFEANTGTDDYGGDSQFNKRAYVGGSDVFQNEKNKLLOSPPKQIASLANGALGI 301
Db 258 HKYLVFEANTGTENYQGEESLFNKAYVGGGTFFRKSQKIQQSARKKDAELANGALGI 317
QY 302 VELADDDYVKSVMKPLVASNTVADEVERANIFKNNKWTLPFTDSGSKMTSDGINDKDYV 361
Db 318 IELNDYTLTKKVMKPLITSNTVTDEIERANVFKMKGWLTFTDSGSKMTIDGINSNDIY 377
QY 362 MLCPGGDSLNGPHNPINETGLVLMNLDLPADLTHTYSHCGIPHPGNNVLTSTYTNNGF 421
Db 378 MLCYVNSLTPGPKYKLNKTGLVLMGLODPNDVTFYSHFAVQAQGNVITSTYTNNGF 437
QY 422 YPEHSHLRDKLGVNKGSDTSGGENS-SGQGG 453
Db 438 FEDKKATFAPSFLMNKGNKTSYVKNLSILEQGG 470

RESULT 6

ABG11754 standard; Protein: 774 AA.

ABG11754;

18-FEB-2002 (first entry)

Novel human diagnostic protein #11745.

Human: chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

W0200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI: 2001-639362/73.

N-PSDB: AAS75941.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity

Claim 20; SEQ ID No 42113; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in CC diagnostics, forensics, gene mapping, identification of mutations CC responsible for genetic disorders or other traits to assess biodiversity CC and to produce other types of data and products dependent on DNA and CC amino acid sequences. ABG00010-ABG3037 represent novel human CC diagnostic amino acid sequences of the invention. CC Note: The sequence data for this patent did not appear in the printed CC specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 774 AA.

Query Match 63.3%; Score 1553; DB 22; Length 774;

Best Local Similarity 65.3%; Pred. No. 4.4e-121;

Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;

QY 2 NSGDYKEDYGFPHITRADMLKIPGQONSQFQVPPQPMASATINDSACYDKSGNLIDD 61

Db 35 NOKAYKETYGVSHITRHDMLQIPKQOQNEKYVPOFQDSTIKNISAKG-----ID 85

QY 62 VMDSWPLOWADGTVAEYNGYHVSALAGDPKNSDDPPLHLEFYQKYGDSIDSMKNAGRVF 121

Db 86 VMDSWPLOWADGTVAEYNGYHVSALAGDPKNSDDPPLHLEFYQKYGDSIDSMKNAGRVF 145

QY 122 EDMKFPVNDPYLKYQTOEMSGSATLTKDQVRLFTYDYGNDPEDGATGAGNIIISTAQV 181

Db 146 KDSKFDANDPLIKDQTOEMSGSATFTSDGKIRLFYTDYSGK -----HYGQSILTTAQV 199

QY 182 NLSQPDATLTKVDGYSVDRKSVFDDGDTYVONIQQFIDEGKWIISGDNHTLRDPHYVEDKG 241

Db 200 NVSKSD-DTLKINGVEDHKTIFD-GDGKTYQVQVOQFIDEGANTSDNHTLRDPHYVEDKG 257

QY 242 HKYLVFEANTGTDDYGGDSQFNKRAYVGGSDVFQNEKNKLLOSPPKQIASLANGALGI 301

Db 258 HKYLVFEANTGTENYQGEESLFNKAYVGGGTFFRKSQKIQQSARKKDAELANGALGI 317

QY 302 VELADDDYVKSVMKPLVASNTVADEVERANIFKNNKWTLPFTDSGSKMTSDGINDKDYV 361

Db 318 IELNDYTLTKKVMKPLITSNTVTDEIERANVFKMKGWLTFTDSGSKMTIDGINSNDIY 377

QY 362 MLCPGGDSLNGPHNPINETGLVLMNLDLPADLTHTYSHCGIPHPGNNVLTSTYTNNGF 421

Db 378 MLCYVNSLTPGPKYKLNKTGLVLMGLODPNDVTFYSHFAVQAQGNVITSTYTNNGF 437

QY 422 YPEHSHLRDKLGVNKGSDTSGGENS-SGQGG 453

Db 438 FEDKKATFAPSFLMNKGNKTSYVKNLSILEQGG 470

RESULT 7

ABG04946 standard; Protein: 789 AA.

ABG04946;

13-FEB-2002 (first entry)

Novel human diagnostic protein #4937.

Human: chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

W0200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.


```

Db      401 VWDSPWLNADGTAAYNGYHVFALAGSPKADADTSLYMFYQKVGDSIDSMKNAGRVE
QY      122 EDMKFVNDPYLKYQTOEMSGSATLTGDOVRLFTYTSNPNEDGGTGAGAGQIISTAYOV
Db      461 KDSKFEFANDPILDKQTOEMSGSATFTSDGKIRLFYTDYSGK-----HYGKOSLTTAYOV
QY      182 NLSQPDAAITLVKDVSDHKSVEFDGDTGYVONIOQFIDEGKMWISGDNHTLDRPHYVEDKG
Db      515 NVKSSD-DTLKINGVEDHKITFD-GDGKTYQNVQOFIDEGVNTSGDNHTLDRPHYVEDKG
QY      242 HKYLVFEANTGTGTCGQDGSFNNKAYGSDVFFQNKKNLQSPKQIASLANGALGI
Db      573 HKYLVFEANTGTGTCGQDGSFNNKAYGSDVFFQNKKNLQSPKQIASLANGALGI
QY      302 VELADDTYKSVKPLVASNTVADEVERANIFKNNKMYLFTDSRGSKMTSDGINDKDYV
Db      633 IELNNDYTLKKYMKPLITSNTVTDEIERANVFKMNGKMYLFTDSRGSKMTIDGINSNDIY
QY      362 MLGPGGDSLNGPHNINFTGLVLMNLDPADLTHYSHGCIPIHPEGNNVLTSTYTNNGF
Db      693 MLGYVNSLTPGYKPLNKTGLVLQMGDLPNDVTFYSHFAVQAKGNVVTSTYTNNGF
QY      422 YPEHSHLRDKLVGNIKGSDTSGGNS-SGQGO 453
Db      753 FEDKKATFAPSPFLMNKIKGNTSVVKNSTILEQGO 785

RESULT 9
ABG28407
ID      ABG28407 standard; Protein: 789 AA.
AC      ABG28407;
XX      18-FEB-2002 (first entry)
DE      Novel human diagnostic protein #28398.
KW      Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX      food supplement; medical imaging; diagnostic; genetic disorder.
OS      Homo sapiens.
PN      WO200175067-A2.
PD      11-OCT-2001.
PE      30-MAR-2001; 2001WO-US08631.
PR      31-MAR-2000; 2000US-0540217.
XX      23-AUG-2000; 2000US-0649167.
PA      (HYSE-) HYSEQ INC.
XX      Drmanac RT, Liu C, Tang YT;
XX      WPI: 2001-639362/73.
DR      N-PSDB; AAS92594.
XX      New isolated polynucleotide and encoded polypeptides, useful in
PT      diagnostics, forensics, gene mapping, identification of mutations
PT      responsible for genetic disorders or other traits and to assess
PT      biodiversity
PS      Claim 20: SEQ ID NO 58766; 103pp; English.
XX      The invention relates to isolated polynucleotide (I) and
XX      polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX      and gene mapping, and in recombinant production of (II). The
XX      polynucleotides are also used in diagnostics as expressed sequence tags
XX      for identifying expressed genes. (I) is useful in gene therapy techniques
XX      to restore normal activity of (II) or to treat disease states involving

```

```

CC      (II). (II) is useful for generating antibodies against it, detecting or
CC      quantitating a polypeptide in tissue, as molecular weight markers and as
CC      a food supplement. (II) and its binding partners are useful in medical
CC      imaging of sites expressing (II). (I) and (II) are useful for treating
CC      disorders involving aberrant protein expression or biological activity.
CC      The polypeptide and polynucleotide sequences have applications in
CC      diagnostics, forensics, gene mapping, identification of mutations
CC      responsible for genetic disorders or other traits to assess biodiversity
CC      and to produce other types of data and products dependent on DNA and
CC      amino acid sequences. ABG00010-ABG3037 represent novel human
CC      diagnostic amino acid sequences of the invention.
CC      Note: The sequence data for this patent did not appear in the printed
CC      publication, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pcl_sequences.
SQ      Sequence 789 AA:
Query Match      63.3%; Score 1553; DB 22; Length 789;
Best Local Similarity 65.3%; Pred. No. 4,5e-121;
Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;
QY      2 NSGDYKEDYGFATHRADMLKIPGOONSPOEFVPOFNASAIKNIDSARKYDGSNLIOLD
Db      350 NOKATKETYGVSHIRHMLQIPKQOQNEKYVOPFDOSTIKNIESAKG-----LD 400
QY      62 WDSWPLNADGTAANYHGHIVLSALAGDPKNSDPTPLHLYQKYGDTSIDSMKNAGRVE
Db      401 WDSWPLNADGTAAYNGYHVFALAGSPKADADTSLYMFYQKVGDSIDSMKNAGRVE
QY      122 EDMKFVNDPYLKYQTOEMSGSATLTGDOVRLFTYTSNPNEDGGTGAGAGQIISTAYOV
Db      461 KDSKFEFANDPILDKQTOEMSGSATFTSDGKIRLFYTDYSGK-----HYGKOSLTTAYOV
QY      182 NLSQPDAAITLVKDVSDHKSVEFDGDTGYVONIOQFIDEGKMWISGDNHTLDRPHYVEDKG
Db      515 NVKSSD-DTLKINGVEDHKITFD-GDGKTYQNVQOFIDEGVNTSGDNHTLDRPHYVEDKG
QY      242 HKYLVFEANTGTGTCGQDGSFNNKAYGSDVFFQNKKNLQSPKQIASLANGALGI
Db      461 KDSKFEFANDPILDKQTOEMSGSATFTSDGKIRLFYTDYSGK-----HYGKOSLTTAYOV
QY      182 NLSQPDAAITLVKDVSDHKSVEFDGDTGYVONIOQFIDEGKMWISGDNHTLDRPHYVEDKG
Db      515 NVKSSD-DTLKINGVEDHKITFD-GDGKTYQNVQOFIDEGVNTSGDNHTLDRPHYVEDKG
QY      242 HKYLVFEANTGTGTCGQDGSFNNKAYGSDVFFQNKKNLQSPKQIASLANGALGI
Db      573 HKYLVFEANTGTGTCGQDGSFNNKAYGSDVFFQNKKNLQSPKQIASLANGALGI
QY      302 VELADDTYKSVKPLVASNTVADEVERANIFKNNKMYLFTDSRGSKMTSDGINDKDYV
Db      633 IELNNDYTLKKYMKPLITSNTVTDEIERANVFKMNGKMYLFTDSRGSKMTIDGINSNDIY
QY      362 MLGPGGDSLNGPHNINFTGLVLMNLDPADLTHYSHGCIPIHPEGNNVLTSTYTNNGF
Db      693 MLGYVNSLTPGYKPLNKTGLVLQMGDLPNDVTFYSHFAVQAKGNVVTSTYTNNGF
QY      422 YPEHSHLRDKLVGNIKGSDTSGGNS-SGQGO 453
Db      753 FEDKKATFAPSPFLMNKIKGNTSVVKNSTILEQGO 785

RESULT 10
ABG25769
ID      ABG25769 standard; Protein: 823 AA.
AC      ABG25769;
XX      18-FEB-2002 (first entry)
DE      Novel human diagnostic protein #25760.
KW      Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX      food supplement; medical imaging; diagnostic; genetic disorder.
OS      Homo sapiens.
PN      WO200175067-A2.
PD      11-OCT-2001.
XX

```

```

PF 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HSE-) HSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX N-PSDB; AAS89956.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity
XX
XX Claim 20; SEQ ID NO 56128; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 823 AA:
XX
XX Query Match 63.3%; Score 1553; DB 22; Length 823;
XX Best Local Similarity 65.3%; Pred. No. 4.8e-121;
XX Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;
XX
XX 2 NSGDYKEDYGFARHTRAMLKIPGQNSPQFVPPFNFASATINIDSAKCYDKSGMLDLD 61
XX I : I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
XX 384 NOKAVKETTYGVSHIRHMDLQIPKQONEKYOVPOFDOSTIKINISAKG-----LD 434
XX
XX 62 VWDSPPLNADGTANVGHYHVSALADPKNSDTPHLFQOKGDPNSIDSMKNAGRYE 121
XX I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
XX 435 VWDSPPLNADGTAEVNGYHVFALASPKDADDTSTYMFQKGDNSIDSMKNAGRYE 494
XX
XX 122 EDMDKFVNDPYLYKQTOEMSGSATLFRDGOVRLFYTYTSGNPEDGCTGAGNOIISTAOV 181
XX I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
XX 495 KDSDFEANDPILIKQTOEMSGSAFTSDGKIRLFLYTYSGK-----HYGQSLTFAOV 548
XX
XX 182 NLSQPDATILKVDGSDHKSVDGSDGTYYQNIQOIFIDGKWIISDNTLRLPHVYEDKG 241
XX I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
XX 549 NVSKSD-DLTKINGVEDHKITFD-GDGKTYOVWQOFIDEGNVTSDGNTLRLRPHVYEDKG 606
XX
XX 242 HKYLVFEANTGTGTDYOGDSFNNKAYYGGSDVFQNKKNKLOSPKQOIASLANGALGI 301
XX I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
XX Db HKYLVFEANTGTENGYGSEESFNKAYYGGGTFNFKRSOKRQOAKRDELALGALGI 666
XX
XX 302 VELADDTYKSVMKPLVASNTVVADEVERANIFKNNKNWYLFDSGSKMTSDGINDKDYV 361
XX I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
XX Db IELNNDYLYKVMKRLINSNTVTDEIERANVFKMGNKYLLFTDSGSKMTIDGINSNDIY 726
XX
XX 362 MLGREGDSLNPHPNINTGLVNNKNDLPADLTHYISCGGIPHPGNNVLTSTYTNRGF 421
XX I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I

```

```

Db 727 MLGYVNSLTPGEKPLNKTLGLWGLDLPNDVFTTYSFAVPAQKGNVVTSTYTNRGF 786
OY 422 YPEHSHLRDKLGAVNIKSDPTSGCENS-SGOGQ 453
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 787 FEDKATRAPSPFLMNKIKNTSVKNSILEGQ 819
XX
XX RESULT 11
XX ABG04500
XX ID ABG04500 standard; Protein; 855 AA.
XX
XX ABG04500;
XX
XX 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #4491.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX (HSE-) HSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX N-PSDB; AAS68687.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity
XX
XX Claim 20; SEQ ID NO 34859; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 855 AA:
XX
XX Query Match 63.3%; Score 1553; DB 22; Length 855;
XX Best Local Similarity 65.3%; Pred. No. 5.1e-121;
XX Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;
XX

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:04:38 ; Search time 41.0246 Seconds

(Without alignments)
326.327 Million cell updates/sec

Title: US-09-986-682B-3

Perfect score: 2454
Sequence: 1 MNSGDYKEDYGFHAHTRADM.....NIKSGDTSGGENSSGQGGFP 455

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCRTS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2454	100.0	455	2	US-08-870-827-3
2	2454	100.0	455	4	US-09-317-179-3
3	281.5	11.5	543	1	US-08-362-232-2
4	281.5	11.5	543	1	US-08-814-196-2
5	278.5	11.3	578	4	US-09-503-172A-2
6	255	10.4	415	1	US-08-381-936-2
7	255	10.4	415	3	US-08-943-374-2
8	128	5.2	1475	3	US-09-007-999-2
9	128	5.2	1475	4	US-09-210-361-2
10	128	5.2	1475	4	US-09-740-274-2
11	116.5	4.7	1848	4	US-08-296-791-6
12	116.5	4.7	1848	5	PCT-US95-10661A-6
13	116	4.7	513	4	US-09-134-001C-4469
14	115.5	4.7	1545	4	US-08-296-791-4
15	115.5	4.7	1545	5	PCT-US95-10661A-4
16	115	4.7	931	4	US-08-624-655A-2
17	115	4.7	1430	3	US-09-008-172-2
18	115	4.7	1430	4	US-09-210-361-6
19	115	4.7	1430	4	US-09-740-274-6
20	114	4.6	21	2	US-08-870-827-1
21	114	4.6	21	4	US-09-317-179-1
22	112.5	4.5	2314	4	US-09-268-347-49
23	110	4.5	1073	4	US-09-206-942-47
24	110	4.5	1079	4	US-09-206-942-47
25	108	4.4	834	1	US-08-471-033-21
26	108	4.4	834	1	US-08-471-044-21
27	108	4.4	834	2	US-08-463-483A-21

28	108	4.4	834	2	US-08-471-046A-21	Sequence 21, Appl
29	108	4.4	834	2	US-08-470-566B-21	Sequence 21, Appl
30	108	4.4	834	2	US-08-469-334-21	Sequence 21, Appl
31	108	4.4	834	3	US-09-300-529-21	Sequence 21, Appl
32	108	4.4	1382	3	US-09-057-570-2	Sequence 2, Appl
33	108	4.4	1577	2	US-08-793-824-2	Sequence 2, Appl
34	108	4.4	2123	4	US-08-968-685A-10	Sequence 10, Appl
35	107.5	4.4	725	4	US-09-668-113A-8	Sequence 8, Appl
36	107.5	4.4	2048	4	US-09-268-347-48	Sequence 48, Appl
37	107	4.4	1833	4	US-08-621-944A-4	Sequence 4, Appl
38	107	4.4	1833	4	US-08-945-567D-4	Sequence 4, Appl
39	107	4.4	1992	4	US-08-621-944A-3	Sequence 3, Appl
40	107	4.4	1992	4	US-08-945-567D-3	Sequence 3, Appl
41	105.5	4.3	1541	4	US-08-296-791-3	Sequence 3, Appl
42	105.5	4.3	1541	5	PCT-US95-10661A-3	Sequence 3, Appl
43	105	4.3	1657	3	US-09-057-570-2	Sequence 2, Appl
44	105	4.3	1805	3	US-09-057-570-7	Sequence 7, Appl
45	104	4.2	1507	6	5268270-2	Patent No. 5268270

ALIGNMENTS

```
RESULT 1
US-08-870-827-3
; Sequence 3, Application US/08870827
; Patent No. 5962297
; GENERAL INFORMATION:
; APPLICANT: Tsusaki et al.
; TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
; TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,827
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 170,630/1996
; FILING DATE: 10-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI=2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-737-3528
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 455 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-870-827-3

Query Match      100.0%   Score 2454; DB 2; Length 455;
Best local Similarity 100.0%; Pred No. 1.5e+208;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 MNSGDYKEDYGFHAHTRADMLKIPGOONSPOFVPOFNASAKINISAKYDRSGNIDL 60
|||||
DB 1 MNSGDYKEDYGFHAHTRADMLKIPGOONSPOFVPOFNASAKINISAKYDRSGNIDL 60
|||||

THIS PAGE BLANK (USPTO)

QY 61 DWDSMPLONADGTAAHYHIVSALAGDPKNSDDTPLHLYOKYGDTSIDSKNAGRY 120
DB 61 DWDSMPLONADGTAAHYHIVSALAGDPKNSDDTPLHLYOKYGDTSIDSKNAGRY 120
QY 121 FEDMDKFPVNDPYLKYQTOEWSGSATLTKDQVRLFTYDYSNPEDEGTAGAGNOIISTAQ 180
DB 121 FEDMDKFPVNDPYLKYQTOEWSGSATLTKDQVRLFTYDYSNPEDEGTAGAGNOIISTAQ 180
QY 181 VNLSDPDAATLKYDGVSDHKSVPDGGDGTYYONIQQFIDEGKWTISGDNHTLRDPHYVEDK 240
DB 181 VNLSDPDAATLKYDGVSDHKSVPDGGDGTYYONIQQFIDEGKWTISGDNHTLRDPHYVEDK 240
QY 241 GHKYLVEFANTGTDDGYOGDQSFNNKAYYGSDVFEFONEKNKLLQSPKKQIASLANGALG 300
DB 241 GHKYLVEFANTGTDDGYOGDQSFNNKAYYGSDVFEFONEKNKLLQSPKKQIASLANGALG 300
QY 301 IVELADDTYVKSVMPLVASNTVADEVERANIFKMNKKWYLFIDSRSKMTSDGINDKDY 360
DB 301 IVELADDTYVKSVMPLVASNTVADEVERANIFKMNKKWYLFIDSRSKMTSDGINDKDY 360
QY 361 YMLGGGSLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHPGNNVLTSTWTRNG 420
DB 361 YMLGGGSLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHPGNNVLTSTWTRNG 420
QY 421 FYPEHSHLRDKLGVNKGSDTSGGENSSGGOFP 455
DB 421 FYPEHSHLRDKLGVNKGSDTSGGENSSGGOFP 455

RESULT 2
US-09-317-179-3

Sequence 3, Application US/09317179
Patent No. 6383769

GENERAL INFORMATION:

APPLICANT: Tsusaki et al.

TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE ACTIVITY

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/317,179

FILING DATE: 24-May-1999

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/870,827

FILING DATE: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TSUSAKI-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 455 amino acids

TYPE: amino acid

MOLECULE TYPE: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

317-179-3

Query Match 100.0%; Score 2454; DB 4; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.5e-208;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSGDKREDYGFHITRADLKIIPGOONSQFVPPQNASAKINDSAGKDYSGNIDL 60
DB 1 MNSGDKREDYGFHITRADLKIIPGOONSQFVPPQNASAKINDSAGKDYSGNIDL 60
QY 61 DWDSMPLONADGTAAHYHIVSALAGDPKNSDDTPLHLYOKYGDTSIDSKNAGRY 120
DB 61 DWDSMPLONADGTAAHYHIVSALAGDPKNSDDTPLHLYOKYGDTSIDSKNAGRY 120
QY 121 FEDMDKFPVNDPYLKYQTOEWSGSATLTKDQVRLFTYDYSNPEDEGTAGAGNOIISTAQ 180
DB 121 FEDMDKFPVNDPYLKYQTOEWSGSATLTKDQVRLFTYDYSNPEDEGTAGAGNOIISTAQ 180
QY 181 VNLSDPDAATLKYDGVSDHKSVPDGGDGTYYONIQQFIDEGKWTISGDNHTLRDPHYVEDK 240
DB 181 VNLSDPDAATLKYDGVSDHKSVPDGGDGTYYONIQQFIDEGKWTISGDNHTLRDPHYVEDK 240
QY 241 GHKYLVEFANTGTDDGYOGDQSFNNKAYYGSDVFEFONEKNKLLQSPKKQIASLANGALG 300
DB 241 GHKYLVEFANTGTDDGYOGDQSFNNKAYYGSDVFEFONEKNKLLQSPKKQIASLANGALG 300
QY 301 IVELADDTYVKSVMPLVASNTVADEVERANIFKMNKKWYLFIDSRSKMTSDGINDKDY 360
DB 301 IVELADDTYVKSVMPLVASNTVADEVERANIFKMNKKWYLFIDSRSKMTSDGINDKDY 360
QY 361 YMLGGGSLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHPGNNVLTSTWTRNG 420
DB 361 YMLGGGSLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHPGNNVLTSTWTRNG 420
QY 421 FYPEHSHLRDKLGVNKGSDTSGGENSSGGOFP 455
DB 421 FYPEHSHLRDKLGVNKGSDTSGGENSSGGOFP 455

RESULT 3
US-08-362-232-2

Sequence 2, Application US/08362232

Patent No. 5641667

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Fructosyltransferase Enzyme, Method

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann & Baron

STREET: 350 Jericho Turnpike

CITY: Jericho

STATE: New York

COUNTRY: United States of America

ZIP: 11758

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 MB

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Wordperfect 6.0 for DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/362,232

FILING DATE: 22-December-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CG 125/93

FILING DATE: 23-December-1993

ATTORNEY/AGENT INFORMATION:

NAME: Baron, Ronald J.

REGISTRATION NUMBER: 29,281

REFERENCE/DOCKET NUMBER: 294-29

TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 822-3550

TELEFAX: (516) 822-3582

INFORMATION FOR SEQ ID NO: 2:

2

THIS PAGE BLANK (USPTO)

SEQUENCE CHARACTERISTICS:

LENGTH: 543 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOHETICAL: NO
US-08-362-232-2

Query Match 11.5%; Score 281.5; DB 1; Length 543;
Best Local Similarity 26.4%; Pred. No. 1.4e-16;
Matches 120; Conservative 59; Mismatches 184; Indels 91; Gaps 26;

QY 5 DYKEDYGFANITRADMLKI-----PGQNSP-QFKVPQFNAS-AIKNIDSAGYDK 53
DB 38 DPQSDP-TARWTRDALQIKASDATVAAGNSLPALQIMPNIADFPVIN----- 87
QY 54 SGNLIDLVWDSMPLOADGTANYHGYHVSALADPK---NSDDTPLH-----LFYQKV 106
DB 88 ----PDVWVWDVMTLIDKHADQFSYNGWEVIFCLTADPNAGYGFDDRHVARIGFEYRRA 143
QY 107 GDTSI-----DSWKNAGRVFED--MDKFVPNDPYLKYOTQEMSGSATLTK--DGOVRLFY 157
DB 144 GIRASRRPVNGWGTGCHLFPDGASAGVYAGQYTT--NQAEWSSSLMLQIHGNTVSFVY 201
QY 158 TDYSGNPEDGCTGAGN---QIISTAOVNLSDPAATLKYDGVSDHKSVFDGGDGTYYON 213
DB 202 TVAFNRD---ANANNITPQALITQTLGRIHADFNHVMFTGFAHPLQ-PDGVLYON 257
QY 214 IQQFIDEGKWSIGDNHILRDPHYVEDKGH---KYLVEANT-----GTTDYGQGSQSNK 266
DB 238 GAO-----NEFNRPDPTFEDPKHPGVNVMVEGNTAGRGVANTTEADLGR-- 306
QY 267 AYYGSDVFEONEKN-KLLOSPKKQIASLANGALGIVELADDTYK--SVMKPLVASNTV 323
DB 307 -----PNDPNAETLQEVLDGSAVYQKANIOLA-IATDSTLSKWKFLSPILSANCY 355
QY 324 ADEVERANIFKNNKMYLFTDSRGSKMTSDGINDKO-VYMLGPGGDSLNGPHNPIN-ETG 381
DB 356 NDQTERPQVYLHNKGYIFTISHTTFAA-GVDGPDGVY--GVGDGIRSDFGPMNYGSG 412
QY 382 LVLMNMLPADLHTYSHCGIPHEGNNVLTYSY 415
DB 413 LTMG---NPTDLNTAGTDFDPSDQNPRAFQSY 443

RESULT 4

US-08-814-196-2

Sequence 2, Application US/08814196

Patent No. 5731173

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Fructosyltransferase Enzyme, Method

TITLE OF INVENTION: For its production and DNA encoding the Enzyme.

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann & Baron

STREET: 350 Jericho Turnpike

CITY: Jericho

STATE: New York

COUNTRY: United States of America

ZIP: 11758

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk, 3.5 inch, 1.44 MB

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Wordperfect 6.0 for DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/814,196

FILING DATE: 10-MAR-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/362,232

FILING DATE: 22-December-1994

APPLICATION NUMBER: CU 125/93

FILING DATE: 23-December-1993

ATTORNEY/AGENT INFORMATION:

NAME: Baron, Ronald J.

REGISTRATION NUMBER: 29,281

REFERENCE/DOCKET NUMBER: 294-29

TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 822-3550

TELEFAX: (516) 822-3582

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 543 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOHETICAL: NO

US-08-814-196-2

Query Match 11.5%; Score 281.5; DB 1; Length 543;
Best Local Similarity 26.4%; Pred. No. 1.4e-16;
Matches 120; Conservative 59; Mismatches 184; Indels 91; Gaps 26;

QY 5 DYKEDYGFANITRADMLKI-----PGQNSP-QFKVPQFNAS-AIKNIDSAGYDK 53
DB 38 DPQSDP-TARWTRDALQIKASDATVAAGNSLPALQIMPNIADFPVIN----- 87
QY 54 SGNLIDLVWDSMPLOADGTANYHGYHVSALADPK---NSDDTPLH-----LFYQKV 106
DB 88 ----PDVWVWDVMTLIDKHADQFSYNGWEVIFCLTADPNAGYGFDDRHVARIGFEYRRA 143
QY 107 GDTSI-----DSWKNAGRVFED--MDKFVPNDPYLKYOTQEMSGSATLTK--DGOVRLFY 157
DB 144 GIRASRRPVNGWGTGCHLFPDGASAGVYAGQYTT--NQAEWSSSLMLQIHGNTVSFVY 201
QY 158 TDYSGNPEDGCTGAGN---QIISTAOVNLSDPAATLKYDGVSDHKSVFDGGDGTYYON 213
DB 202 TVAFNRD---ANANNITPQALITQTLGRIHADFNHVMFTGFAHPLQ-PDGVLYON 257
QY 214 IQQFIDEGKWSIGDNHILRDPHYVEDKGH---KYLVEANT-----GTTDYGQGSQSNK 266
DB 238 GAO-----NEFNRPDPTFEDPKHPGVNVMVEGNTAGRGVANTTEADLGR-- 306
QY 267 AYYGSDVFEONEKN-KLLOSPKKQIASLANGALGIVELADDTYK--SVMKPLVASNTV 323
DB 307 -----PNDPNAETLQEVLDGSAVYQKANIOLA-IATDSTLSKWKFLSPILSANCY 355
QY 324 ADEVERANIFKNNKMYLFTDSRGSKMTSDGINDKO-VYMLGPGGDSLNGPHNPIN-ETG 381
DB 356 NDQTERPQVYLHNKGYIFTISHTTFAA-GVDGPDGVY--GVGDGIRSDFGPMNYGSG 412
QY 382 LVLMNMLPADLHTYSHCGIPHEGNNVLTYSY 415
DB 413 LTMG---NPTDLNTAGTDFDPSDQNPRAFQSY 443

RESULT 5

US-09-503-172A-2

Sequence 2, Application US/09503172A

Patent No. 6284510

GENERAL INFORMATION:

APPLICANT: ITO, Tetsuya

APPLICANT: FUJITA, Koki

APPLICANT: HARA, Kozo

APPLICANT: TONOKURA, Takashi

APPLICANT: SAKANO, Yoshiyuki

TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE GENE

FILE REFERENCE: 10749-0001-0

CURRENT APPLICATION NUMBER: US/09/503,172A

CURRENT FILING DATE: 2000-02-14

PRIOR APPLICATION NUMBER: JP 160416/1999

PRIOR FILING DATE: 1999-06-08

THIS PAGE BLANK (USPTO)

NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO: 2
LENGTH: 578
TYPE: PRT
ORGANISM: Arthrobacter sp.
US-09-503-172A-2

Query Match 11.3%; Score 278.5; DB 4; Length 578;
Best Local Similarity 23.7%; Pred. No. 2.7e-16;
Matches 125; Conservative 67; Mismatches 193; Indels 143; Gaps 25;

8 EDYGAHTYRADMLK-----GQOQNSPOFK-----VPOFNASAIKIDSARKYDSS 54
57 EDDFAKTRADARLOKSDPTAHSREMSPASTMPVPO----- 98
55 GNLIDL-----VMSWPLQNDGTRANYHGYHVSALAGDKP-NSDD-----TPHLFYQ 104
99 -DFPMSNEQWVWMDTFLTDADANQSVNGWEIIFSLVADRNLGEDRHFVAKIGFYR 157
105 KVGDTSID-----SKKNAGRVT-EDMDKFPVNDPIKYOTQWSSGATLTKGQVRLFT 158
158 PAGVPAERPENGGTYGGLVKEGVTGQIFEDQSFHQTO-WGSGARVSKNGEIKLFT 216
159 DYSGNPEDGCTGA-----GNQISTAQVNLSPDAATLKVGVSDHKSVPDGGGTYQNI 214
217 DVAFNRNDGTNIRKYDRIALSVKRVANKGCVLLTGKNTDLDQ-----ADGITYQIG 272
215 QQFIDEGKWISGDNHTLRDPHYVEDKGH--KYLVFEAN-----TGTTD-----GY-QGD 260
273 AQ-----NEFENFRDPFTEEDPAHGEFTFVVEGNSAMQREIATCNEADLGVRQGD 323
261 QSFNNKAYGSGDVEFQNEKKNLLOSPKKQIASLANGALGYELADDTYKSVMPYVAS 320
324 PYAETVDVNASGATYQIGNGLAKAKNKOL-----TEWEFLPILISA 366
321 NTVADEVERANIFKNNKMYLFTDSRGSKMTSDGINDKDYMLPGGDSLDLPHNIN-E 379
367 NCVIDQTRPOLYFEDGKSYLFTISHRGTFAGLDGPEGVY--GPGGDSIRSDYOLNKG 424
360 TGLVLMNLDPADLT-----HTYSHGCIHPHPEGNNVLTSYMTNRGF 421
425 SGLALG-----NPTNLNLFGLGGPAPDFNQHPGHQAYSHYMP-----GGLVQSFIDTIG- 475
422 YPEHSHLRD-----KLGVNIKSDTS-----GGENSSGQGF 455
476 --THDFVRGGTLAPTVKMDIGVGDPTKTAVDYSYSGELGGWADIP 521

RESULT 6

US-08-381-936-2
Sequence 2, Application US/08381936
Patent No. 5792923
GENERAL INFORMATION:
APPLICANT: ROBER, Manuela
APPLICANT: GEIER, Gebhardt
APPLICANT: GEIDER, Klaus
TITLE OF INVENTION: DNA sequences which lead to the
TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing
TITLE OF INVENTION: these sequences as well as a process for preparing
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/381,936

FILING DATE: 09-FEB-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 93 02110

FILING DATE: 09-AUG-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P42270618

FILING DATE: 08-DEC-1992

ATTORNEY/AGENT INFORMATION:
NAME: Mellman, Edward A.

REGISTRATION NUMBER: 24,735

REFERENCE/DOCKET NUMBER: P/951-108

TELEPHONE: (212) 382-0700

TELEFAX: (212) 382-0888

TELEX: 236925

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 415 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-381-936-2

Query Match 10.4%; Score 255; DB 1; Length 415;
Best Local Similarity 25.4%; Pred. No. 2e-14;

Matches 115; Conservative 64; Mismatches 141; Indels 132; Gaps 29;

9 DYGFHAI--TRADMLKIPGQNSPOFKVPOFNASAIKIDSARKYDSSGMLIDL----- 61
3 DYVKKFTLWTRADALRV--HEDDPTTTP-----VIDIAPVMS 39
62 -----VMSWPLQNDGTRANYHGYHVSALAGDKPNSDDTPLH----- 100
40 EEVFTIDTMLRPFDEIISVNGWCITFTLTAD-RNTDNEQFODENGNNITRDWEDRHG 98
101 -----LFYQKVGDTSIDSKNAGRVEEDMDKFPVNDPIKYOTQWSSGATLTKD-GQVR 154
99 RAICVWYSRTG-----KDMIFGRVM--AEGVAP-----TRENAGITLLNDRDID 145
155 LEYTDYSGNPEDGCTGA-----GNQISTAQVNLSPDAATLKVGVSDHKSVPDGGGTYV 211
146 LYTTCVT-----PGATIAKVRGKIVTSDQSVSL-----BGFOOVTSLF-SADGTY 190
212 QNIQCFIDEGKWTSGNHTLRDPHYVEDK--GHKYLVFEANTGTTCGTCGQDQSFNNKAY 269
191 QTEEQ--NAFW-----NFRDPSPIIDRDGKLYMLFEGNVA--GPRQSHET-TOAEM 237
270 GGSDFVQNEKKNLLOSPKKQIASLANGALGYELADDTYK--SVKKPLVASNTVADEV 327
238 GNVPCIEYDVGAKKYA-----GCYGLA-VAKDLSGSEMOLIPLTITAVGVNDOT 286
328 ERANIFKNNKMYLFTDSRGSKMT-SDGINDKD-VYMLGPGGDSLNGPHNINETGLVLT- 384
287 ERHEHFEQGGKXYLFTISH--KYTFADNLTGPGGVY--GVSQKLGKPYTPMSSGLVIG 342
385 NMNLDPADLTHYSHGCIHPHPEGNNVLTYSYM 416
343 NPSSOP--FOTYSHYMP-----NGLVTSFI 366

RESULT 7

US-08-943-374-2
Sequence 2, Application US/08943374
Patent No. 6028249
GENERAL INFORMATION:
APPLICANT: ROBER, Manuela
APPLICANT: GEIER, Gebhardt
APPLICANT: GEIDER, Klaus

THIS PAGE BLANK (USPTO)

APPLICANT: WILLMITZER, Lothar
TITLE OF INVENTION: DNA sequences which lead to the
formation of polyfructans (levans), plasmids containing
TITLE OF INVENTION: these sequences as well as a process for preparing
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ostrolenk, Faber, Gerb & Stoffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,374
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/381,936
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P42270618
FILING DATE: 08-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Mellman, Edward A.
REGISTRATION NUMBER: 24,735
REFERENCE/DOCKET NUMBER: P/951-108
COMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 415 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-374-2

Query Match 10.4%; Score 255; DB 3; Length 415;
Best Local Similarity 25.4%; Pred. No. 2e-14;
Matches 115; Conservative 64; Mismatches 141; Indels 132; Gaps 29;

QY 9 DYGFANI--TRADMLKIPGOONSPOFKYPOFNASAIKNIDSAGKYDKSGNLLID-- 61
DB 3 DYVKKPTLMTTRADALKV--HEDPTTTP--VIDIAFPVMS 39
QY 62 ---VWDSMPLQADGTAANYHGYHYVSALAGPKNSDPTLH-- 100
DB 40 EEFYVDMPLRDEDEIISVNCWCIIETLAD-RMTDNPQFODENGNYDITRDWEDRHG 98
QY 101 ---LEFYKVPDTSIDSMKNAGRVEDMDKFEVNDPYLKYQOEWGSGTLTKD--GQVR 134
DB 99 RARICWYSRTG---KDMIFGSRV--AEGVAPT---FREAAGTPTLLINDRID 145
QY 155 LFYTYDSGNPEDGGTGA---GNOIISTAVQNLSPDPAATLVGVSDHKVSPFGGGGTYY 211
DB 146 LYTTCTVT---PGATIAKVRGKIIVTSDQSVSL---EGFOQVTSLF--SADGTIY 190
QY 212 QNIOGFIDGKWIISGDNHTLRDPHYEDK--GHKTYLVEANNTGTTDGYGDSFNKKAYY 269
DB 191 QTEEO---NAFW---NFRDPSPEIDRNDGKLYMLFEGNVA---GPRGSHET--TQAEK 237
QY 270 GSDVFEQNEKKNKLLQSLASLANGALGIVEADDTYK--SYMKPLVASNTVADEVY 327
DB 238 GNVPGYDEYVGAKYQA---GCVGLA--VAKDLSGSEWQTLPLPLTAIVGVNDQT 286
QY 328 ERANIFKNNKWLFTDSRGSKMT--SDGINDKD--VYMLPGGDSLGNPHNPINETGLV-- 384

DB 287 ERPHFVQDGKXYLFTTISH--KYTFADNLTPDGVY--GFVSDMLTGTYTMSNGVLVG 342
QY 385 NMNIDPADLFTYTHCGIPHEGNNVLTSM 416
DB 343 NPSSQP---FQYTSHYVP---NGLVTSPI 366

RESULT 8
US-09-007-999-2
Sequence 2, Application US/09007999
Patent No. 6087559
GENERAL INFORMATION:
APPLICANT: Nichols, Scott E.
TITLE OF INVENTION: Substitutes for Modified Starch and
TITLE OF INVENTION: Latexes in Paper Manufacture
FILE REFERENCE: 0356D
CURRENT APPLICATION NUMBER: US/09/007,999
CURRENT FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/478,704
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1475
TYPE: PR1
ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match 5.2%; Score 128; DB 3; Length 1475;
Best Local Similarity 20.4%; Pred. No. 0.022;
Matches 79; Conservative 48; Mismatches 102; Indels 158; Gaps 22;

QY 68 LONADGTANYH--YHVSALADDPKNSDPTLHLPYQKYGDSISWKNA 117
DB 1131 LKNEDGTYA--YIGNDGRRENGYQFMGWRHFNNGE---MSVGLTVID-- 1176
QY 118 GRV--FEDMDKFPNDPYLKYQTOEMSGSATLTKDGOVRLFTDYSGN-- 163
DB 1177 GOVYFEDMG---YQAK---GKFTYTAGKIR--YFDKSGNNYRRTIENEG 1222
QY 164 ---PEGGTGAGNOIISTQVNLSPDPAATLVKDYVSDHK---SVEPDGGTYVONI 215
DB 1223 KWLVEGGAANTVSSQITN--QOHLFRANGVQVKEFTDHHGRISYDNGSDQIRN-- 1279
QY 216 QFI--DECKWISGDNHTLRDPHYEDKGH--KLVFEAN-- 250
DB 1280 RFVNNAGQMF---YFDNNGYAVTGARTINGQLLYFRANGVQVKEFTDRIY 1328
QY 251 --TGTTDGYGQD---SFNNKAY---YGGSDVFEQNEKKNKLLQSP 287
DB 1329 GRISYDNGSDQIRNFRVRAQGMWYFDNNGYAVTGARTINGQLLYFR-- 1378
QY 288 KQQLASLANGALGIVEADDTYKSYMKPLVASNTVADEVERANIFKNNKWLFTDSRG 347
DB 1379 ---ANG---VQYKGEFVTRHGRISYDNGSDQIRNFRVRAQGMWYF--DNNG 1426
QY 348 SKMTSDGINDKDVYMLPGGDSLGNPH 374
DB 1427 YAVT---GARTINGOH 1439

RESULT 9
US-09-210-361-2
Sequence 2, Application US/09210361
Patent No. 6284479
GENERAL INFORMATION:
APPLICANT: Nichols, Scott E.
TITLE OF INVENTION: Substitutes for Modified Starches and
TITLE OF INVENTION: Latexes in Paper Manufacture
FILE REFERENCE: 0357CR
CURRENT APPLICATION NUMBER: US/09/210,361
CURRENT FILING DATE: 1998-12-11

THIS PAGE BLANK (USFTO)

EARLIER APPLICATION NUMBER: 09/007,999
EARLIER FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/478,704
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 09/009,620
EARLIER FILING DATE: 1998-01-20
EARLIER APPLICATION NUMBER: 08/485,243
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 09/008,172
EARLIER FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/482,711
EARLIER FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
FASTSEQ for Windows Version 3.0
LENGTH: 1475
TYPE: PRF
ORGANISM: Streptococcus mutans
US-09-210-361-2

Query Match
Best Local Similarity 5.2%; Score 128; DB 4; Length 1475;
Pred. No. 0.022;
Matches 79; Conservative 48; Mismatches 102; Indels 158; Gaps 22;

68 LONADGTANYHG-----YHVSALAGDPKNSDTPHLFYOKVGTSTDSWKNA 117
1131 LKNEDETYA-YGNDGRRENGYQFMGVRHFNNGE-----MSVGLTVID----- 1176
118 GRV-FEDMDKFEVNDPYLKYOTQEMSGATLTKDQGVRLFTYDYSN----- 163
1177 GQVOYFDEM-----YQAK---GKFTVTADGKIR-YFDKQSGMTRNRFIENEG 1222
1177 GQVOYFDEM-----YQAK---GKFTVTADGKIR-YFDKQSGMTRNRFIENEG 1222
164 -----PEDGTGAGNOIISTAOVNLSDPAATLKVDVSDHK---SVFDGDDGTVYQNIQ 215
1223 KWLYLEDEGAAYVSGTIN-GQHLFRANGVOYKGEFTVDHGRIRISTYDNGSGDQIRN-- 1279
216 QFI-DEGWISGDNHTLRPHYVEDKGH-----KYLVEAN----- 250
1280 RFRVNAOGQMF-----YFDNNGYAVTGARTINGQLLYFRANGVOYKGEFTVDYR 1328
251 -TGTTDGYOGDO-----SFNNKAY-----YGSDFPQNEKNKLLQSP 287
1329 GRISYDNGSGDQIRNFRVNAOGQMFYFDNNGYAVTGARTINGQHLHYER----- 1378
288 KKOIASLANGALGIVELADYTVKSVKPLVASNTVADEVERANIRKMNKNWLTFTDSRG 347
1379 -----ANG-----VOYKGEFTVDHGRIRISTYDNGSGDQIRNFRVNAOGQMFYF-DNNG 1426
348 SKMTSDGINDKDYVYMLGPGDGLNGPH 374
1427 YAVT-----GARTINGQH 1439

RESULT 10
US-09-740-274-2
Sequence 2, Application US-09-740-274
Patent No. 6465203
GENERAL INFORMATION:
APPLICANT: Nichols, Scott E.
TITLE OF INVENTION: Glucan-containing compositions and paper
FILE REFERENCE: 0357CRD
CURRENT APPLICATION NUMBER: US/09/740,274
PRIOR APPLICATION NUMBER: 09/210,361
PRIOR FILING DATE: 1998-12-11
PRIOR APPLICATION NUMBER: 09/007,999
PRIOR FILING DATE: 1998-01-16
PRIOR APPLICATION NUMBER: 08/478,704
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: 09/009,620
PRIOR FILING DATE: 1998-01-20
PRIOR APPLICATION NUMBER: 08/485,243
PRIOR FILING DATE: 1995-06-07

PRIOR APPLICATION NUMBER: 09/008,172
PRIOR FILING DATE: 1998-01-16
PRIOR APPLICATION NUMBER: 08/482,711
PRIOR FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
FASTSEQ for Windows Version 3.0
LENGTH: 1475
TYPE: PRF
ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match
Best Local Similarity 5.2%; Score 128; DB 4; Length 1475;
Pred. No. 0.022;
Matches 79; Conservative 48; Mismatches 102; Indels 158; Gaps 22;

68 LONADGTANYHG-----YHVSALAGDPKNSDTPHLFYOKVGTSTDSWKNA 117
1131 LKNEDETYA-YGNDGRRENGYQFMGVRHFNNGE-----MSVGLTVID----- 1176
118 GRV-FEDMDKFEVNDPYLKYOTQEMSGATLTKDQGVRLFTYDYSN----- 163
1177 GQVOYFDEM-----YQAK---GKFTVTADGKIR-YFDKQSGMTRNRFIENEG 1222
164 -----PEDGTGAGNOIISTAOVNLSDPAATLKVDVSDHK---SVFDGDDGTVYQNIQ 215
1223 KWLYLEDEGAAYVSGTIN-GQHLFRANGVOYKGEFTVDHGRIRISTYDNGSGDQIRN-- 1279
216 QFI-DEGWISGDNHTLRPHYVEDKGH-----KYLVEAN----- 250
1280 RFRVNAOGQMF-----YFDNNGYAVTGARTINGQLLYFRANGVOYKGEFTVDYR 1328
251 -TGTTDGYOGDO-----SFNNKAY-----YGSDFPQNEKNKLLQSP 287
1329 GRISYDNGSGDQIRNFRVNAOGQMFYFDNNGYAVTGARTINGQHLHYER----- 1378
288 KKOIASLANGALGIVELADYTVKSVKPLVASNTVADEVERANIRKMNKNWLTFTDSRG 347
1379 -----ANG-----VOYKGEFTVDHGRIRISTYDNGSGDQIRNFRVNAOGQMFYF-DNNG 1426
348 SKMTSDGINDKDYVYMLGPGDGLNGPH 374
1427 YAVT-----GARTINGQH 1439

RESULT 11
US-08-296-791-6
Sequence 6, Application US/08296791
Patent No. 6245337
GENERAL INFORMATION:
APPLICANT: St. Geme III, Joseph W.
APPLICANT: Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehm, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.

THIS PAGE BLANK (USPTO)

```

;
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELE: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
;
US-08-296-791-6

Query Match          4.7%; Score 116.5; DB 4; Length 1848;
Best Local Similarity 20.8%; Pred. No. 0.32;
Matches 105; Conservative 64; Mismatches 181; Indels 155; Gaps 28;

OY 25 GQONSPOKVPQFNAS--AIKNIDSAGYDKSGNLIIDVWDSWPLQNAADGTAAANYHGYH 82
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 GDSGSPLEFYDREKGMFLGSDYDFWAGYNNKS-----WQEMNIYKHEFAEKIYQOYS 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 83 IVSAL-----AGDPKNSD-----DTPLH----- 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 AGSLIGSNQYTWQATGSTTTGGGEP LSVLDLIDGKKPNNHKSITLKGSGTLFLNNHI 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 101 -----LF-----YOKVGDTSIDSMKNAGRVFEEDMKFYP--NDPYLKYQTQEMSGSATL 147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 DQAGAGLFPEGDYEVKGTSDSTTWKAGVSVAD-GKTVWVKNHP--KYDRLAKIGKGTL 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 148 TKDGOVRLFYTYDSGNP-----DG-----GTGAGNOIISTAQVNL-SQPPAATLKVD 194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 VVEGR-----GKNGLLKVGDGYVILKOKADANNKVOAFSGVIGSGSTLVLLND 516
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 195 GVSDDKSVFDG-----GDGTYVQNTQGFIDGKMTSGDNHRLRDPHYEDGKHLYL 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 KOYDPSNITFGRGRLDNGSLTFDHIRN-IDGARVY--NHNMTNTSNTITGES-L 572
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 246 VFEANTGTIDGYQGDOS-----FN--NKAYYGGSDVFFONEKNKLLQSPK 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 573 ITNPMTITSYNIEAODDHPRLIRSIPIRQLYFNQDNRSY-----TLKGASTRSEL 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 289 KOIASLANGALGIVELADDTYKSVAKPLVASNTVADEVERANIFKNNKWLFTDSRGS 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 PONSSESNMVLVMGRTSDAAKRNNVNI-----NNERNNGF--NGYGESETKAT 674
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 349 K-----MTSDGINDKDVYMLGPGGSLNGPHNPINETGLVLMNIDPADLFTYSHCGI- 402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 QNGKLVTFNGKSDONRFL-TGGTNLNGDLN-VEKGTFLFSGRPP-----HARDIAGIS 728
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 403 -----PH-PEGNNVLTSYMTNRGF 421
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 729 STKKDPHFTENNEVVEEDWIMRNF 753
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
PCT-US95-10661A-6
; Sequence 6, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecautid, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELE: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
;
PCT-US95-10661A-6

Query Match          4.7%; Score 116.5; DB 5; Length 1848;
Best Local Similarity 20.8%; Pred. No. 0.32;
Matches 105; Conservative 64; Mismatches 181; Indels 155; Gaps 28;

OY 25 GQONSPOKVPQFNAS--AIKNIDSAGYDKSGNLIIDVWDSWPLQNAADGTAAANYHGYH 82
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 GDSGSPLEFYDREKGMFLGSDYDFWAGYNNKS-----WQEMNIYKHEFAEKIYQOYS 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 83 IVSAL-----AGDPKNSD-----DTPLH----- 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 AGSLIGSNQYTWQATGSTTTGGGEP LSVLDLIDGKKPNNHKSITLKGSGTLFLNNHI 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 101 -----LF-----YOKVGDTSIDSMKNAGRVFEEDMKFYP--NDPYLKYQTQEMSGSATL 147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 DQAGAGLFPEGDYEVKGTSDSTTWKAGVSVAD-GKTVWVKNHP--KYDRLAKIGKGTL 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 148 TKDGOVRLFYTYDSGNP-----DG-----GTGAGNOIISTAQVNL-SQPPAATLKVD 194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 VVEGR-----GKNGLLKVGDGYVILKOKADANNKVOAFSGVIGSGSTLVLLND 516
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 195 GVSDDKSVFDG-----GDGTYVQNTQGFIDGKMTSGDNHRLRDPHYEDGKHLYL 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 KOYDPSNITFGRGRLDNGSLTFDHIRN-IDGARVY--NHNMTNTSNTITGES-L 572
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 246 VFEANTGTIDGYQGDOS-----FN--NKAYYGGSDVFFONEKNKLLQSPK 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 573 ITNPMTITSYNIEAODDHPRLIRSIPIRQLYFNQDNRSY-----TLKGASTRSEL 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 289 KOIASLANGALGIVELADDTYKSVAKPLVASNTVADEVERANIFKNNKWLFTDSRGS 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 PONSSESNMVLVMGRTSDAAKRNNVNI-----NNERNNGF--NGYGESETKAT 674
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 349 K-----MTSDGINDKDVYMLGPGGSLNGPHNPINETGLVLMNIDPADLFTYSHCGI- 402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 QNGKLVTFNGKSDONRFL-TGGTNLNGDLN-VEKGTFLFSGRPP-----HARDIAGIS 728
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 403 -----PH-PEGNNVLTSYMTNRGF 421
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 729 STKKDPHFTENNEVVEEDWIMRNF 753
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-09-134-001C-4469
; Sequence 4469, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
```

CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4469
LENGTH: 513
TYPE: PRF
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4469

Query Match 4.7%; Score 116; DB 4; Length 513;
Best Local Similarity 20.2%; Pred. No: 0.051;
Matches 96; Conservative 62; Mismatches 144; Indels 174; Gaps 22;

QY 86 ALADPKNSDDTPLHLFYQKVGDTSIDSMKNAHVEEDMDKFPNDPYLKYQTEWMSGSA 145
D 57 ALKELPK-SENKKHKYDKYKVTDEK-----NKGFHTTLOPKVGN- 97
QY 146 TLTDGQVRLFTYYSNP--EDGGTGAAGNOIISTAOYNLSQPDAA----- 189
D 98 TYADPKKVKV-HTNKKESKVLVNGDTDA-KKVOPTNKVSIKESATDCAFEAIKIDROKA 155
QY 190 -TLKVDGVSDFKSYFDGDDGVYQNIQ-----OFIDE----- 220
D 156 KNLKSDVYKTKKVELDEKKNYVNIETITSPKISHNNKVIDAETGQVVKLNMKEEA 215
QY 221 ---GKWSIGDN-----TLRDPHYEDKGHKLYFEANTGTDD 256
D 216 TTGKGKGLVDTKQININSVSGYALODLTQGTLSAYNYDANTGOAVLMDKDRNFDD 275
QY 257 YQ-----GDSTNNKAYYGGSVF-----FQNEKNKLLOSPK 288
D 276 EORAGVDANYAKETDYIKNTFFRESTDNO---GSPILSLAHVNNFGODNR----- 325
QY 289 KOIASLANGA--LGIVELADDTYKSVKPLVASTVADDEV-----ERANIFKNNKMY 340
D 326 ---NNAMIDDKMIGDGDRTFTALSGANDVVAHELTHTGTOOTALVYRSOSGA 378
QY 341 L---FTDSRGSKM-TSDGINDKDYVMGLPGDSDLNGPNNPINEGVLVNNMLDPADLTHT 396
D 379 LNESFSDVFGYVDEDFLMEDEYVTPGVGDALRSMSNP--ERFGQPSHMN---DEVT 433
QY 397 ---YSHCGIPHEGNNVY-----LTSYMTNGEYFEHSHLR 430
D 434 NSDNGVHTNSGIRKKAAYNTIRSIGKORSEQIYRALTLYTUSNDFODAKASIQ 489

RESULT 14
US-08-296-791-4
Sequence 4, Application US/08296791
Patent No. 6245337

GENERAL INFORMATION:
APPLICANT: St. Geme III, Joseph W.
APPLICANT: Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Trecautin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1545 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-296-791-4

Query Match 4.7%; Score 115.5; DB 4; Length 1545;
Best Local Similarity 20.8%; Pred. No: 0.3;
Matches 104; Conservative 55; Mismatches 193; Indels 149; Gaps 23;

QY 25 GQNSPOFKYQENAS--AIKNIDSAGYDKSGNMLDLVDWSDWPLONADSTANYHGYH 82
D 290 GDSSGPLEFYDREKGMFLFSGDYDMAGYNKS-----WQENNIYKPEFAEKTYEYS 342
QY 83 IVSALADGPKNSDDTPLHLFYQKVGDTSIDSMKNAH-----VEEDMDKVF 128
D 343 AGSL-----IGSKTDVSWSSNGKSTITTGGEKSLANDLAGKDR-- 381
QY 129 PNDPYLKYQTEWMSGSATLTKD---GOVRLFYTYSGNPEDGCT-----GAGNOIIST 178
D 382 PN-HGKSVFEFGSLTLNNNIDGAGGLE---FEGDYEVKGTSDMTTWKAGVSAEG 436
QY 179 AQV-----NLSDPDAATLVKVDGVSDFKSYFDGDDGVYQNIQOFIDEKMW----- 223
D 437 KTVWVKVNPQYDLIAKIGKGLIVEGTGDKGSLKVGDDGVILK--OOTNGSGHAFASV 495
QY 224 ---ISG-DNHTLRDPHYE-----DKGHKLYFEANTG 252
D 496 GIVSGRSTLVLDNDKQVDPNSTYFGFRGRLDLNGSLTPEHINIDEAGALVNHSHSKH 555
QY 253 TTGDEYGDQSF--NNKAYV-----GGSDFVFEQNEKNKLLOSPKQIA 292
D 556 STVITGDNLTIDPNNNSIYVYKPLEDDNPYAIQIKKYQLYP--NEENPTYALAKDAS 614
QY 293 SLANGALGIVLADDTYKSVKPLVASTVADDEVANIRKMNKMYLFTDSRGS----- 348
D 615 IRSEFPQNRGSSNNSWLYMTEKADAQNA---MNHINERNMGFPNGYFEEEGKNGN 670
QY 349 \KMTSDGINDKDYVMGLPGDSDLNGPNNPINEGVLVNNMLDPADLTHTYSHCGIP----- 403
D 671 LNVTFKKSSEQRNRL-TGNTNLNGDLN-VQGTGLFLSGRTP---HARDIAGISSTKK 724
QY 404 ---HPEGNNVLYSYMTNRGF 421
D 725 DSHFSENNEVVVEDMDIRNF 745

RESULT 15
PCT-US95-10661A-4
Sequence 4, Application PC/TUS9510661A
GENERAL INFORMATION:
APPLICANT: Washington University, et al.
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10661A
FILING DATE: 16-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: PP-59941/REF
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1545 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-4

Query Match 4.7%; Score 115.5; DB 5; Length 1545;

Best Local Similarity 20.8%; Pred. No. 0.3;
Matches 104; Conservative 55; Mismatches 193; Indels 149; Gaps 23;

QY 25 GQONSPOFKYPOFVAS--AIKNIDSAGYDKSGNLIDLDVWDSWPLQADGTANYHGYH 82
DB 290 GDGSGPLFVYDREKGLPLGSIYMAGYNKS-----WQEWNIYKPEFAEKIYECYS 342
QY 83 IVSALAGDPKNSDDTPLHLFFQKYVDTSIDSKNAGR-----VFEDMDKFV 128
DB 343 AGSL-----IGSKTDYSWSSNGKSTITGGEKSLNVDLADGDKD-- 381
QY 129 PNDPLKYOQEWGSGATLTKD---GQVRLFTYDSGNPEDEGT-----GAGNQIIST 178
DB 382 PN--HGKSVTFEGSGTLTLNNNIDQAGGLE--FEGDYEVKGTSDNTYMGAGVSAEG 436
QY 179 AQV-----NLSQPDATLKVDSVSHKSVFSGDGTVYQNIQOFTDEGK----- 223
DB 437 KTYVMKVNPOYDLAKIGKGLIVEGTGDKGSLKYGDGTVILK-QQTNGSGQHAFASV 495
QY 224 --ISG-DNHTLRDPHYE-----DKGHRVYLFVEANTG 252
DB 496 GIVSGRSTLVLDNDKQVDPNSIYFGFRGLDLNGNSLTFDHIRNIDEGARLVNHSTSKH 555
QY 253 TTDDYQGDQSF--NKNAY-----GSDVFFQNEKNLLOSPKQIA 292
DB 556 STVTITDNLITDPNNYSIYVKPLEDDNPYAIRQIKYQOLYF-NEENPTYVALKRDAS 614
QY 293 SLANGALGIVELDADYVKSVMKPLVASNTVADEVERANIFKMNKKWYLFDSRGS--- 348
DB 615 IRSFEPQNRGESNNSWLYMTEKADAKNA---MNIINERMRNGFNGEEDGKNGN 670
QY 349 -KMTSDGINDKDYVMLGPGDLSLNGPHNPINETGLVLMNLDPADLTHYSHGIP--- 403
DB 671 LNVTFKCKSEQNRL-L-TGGTNLNGDLN-VQOGTLFLSGRTP---HARDIAGISSPKK 724
QY 404 ---HPEGNNVYVLSYMTNRGF 421
DB 725 DSHFSENNEVYVEDDWINRNF 745

Search completed: June 16, 2003, 12:11:35
Job time : 44.0246 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:07:33 ; Search time 69.9283 seconds
(without alignments)
695.665 Million cell updates/sec

Title: US-09-986-682B-3
Perfect score: 2454
Sequence: 1 MNSGDYKEDYGFRAHITRADM.....NIKSGDTSGENSSGOGFP 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2454	100.0	455	10	US-09-986-682B-3 Sequence 3, Appl1
2	649.5	26.5	792	10	US-09-995-587A-11 Sequence 11, Appl1
3	606.5	24.7	789	10	US-09-995-587A-1 Sequence 1, Appl1
4	129.5	5.3	1411	9	US-10-080-505-17 Sequence 17, Appl1
5	128	5.2	1475	10	US-09-740-274-2 Sequence 2, Appl1
6	126	5.1	2076	10	US-09-815-242-5815 Sequence 3815, Ap
7	126	5.1	2186	10	US-09-815-242-12913 Sequence 38, Appl1
8	125	5.1	1981	9	US-09-928-457-38 Sequence 1, Appl1
9	125	5.1	2015	9	US-10-066-551-1 Sequence 8, Appl1
10	120	4.9	773	12	US-10-067-385-8 Sequence 6, Appl1
11	116.5	4.7	1848	9	US-10-080-505-6 Sequence 6, Appl1
12	116.5	4.7	1545	9	US-09-839-996-4 Sequence 4, Appl1
13	115.5	4.7	1545	9	US-10-080-505-4 Sequence 381, App
14	115.5	4.7	1430	10	US-09-912-020-381 Sequence 6, Appl1
15	115	4.7	1430	10	US-09-740-274-6 Sequence 9, Appl1
16	114.5	4.7	1434	9	US-10-080-505-9 Sequence 1, Appl1
17	114.5	4.6	21	10	US-09-986-682B-1 Sequence 304, App
18	114	4.6	1325	10	US-09-741-669-304 Sequence 13, Appl1
19	113	4.6	1325	10	US-09-741-669-304 Sequence 5510, Ap

20	113	4.6	1436	9	US-10-080-505-13 Sequence 5510, Ap
21	110.5	4.5	546	10	US-09-815-242-5510 Sequence 12592, A
22	110.5	4.5	549	10	US-09-815-242-12592 Sequence 6, Appl1
23	108	4.4	915	10	US-09-817-514A-6 Sequence 5456, Ap
24	108	4.4	978	10	US-09-815-242-5456 Sequence 12686, A
25	108	4.4	1001	10	US-09-815-242-12686 Sequence 9, Appl1
26	108	4.4	2122	9	US-09-813-214A-9 Sequence 430, App
27	106.5	4.3	650	10	US-09-801-368-430 Sequence 7, Appl1
28	106.5	4.3	2059	9	US-10-192-584-7 Sequence 3, Appl1
29	106	4.3	2516	10	US-09-817-514A-2 Sequence 3, Appl1
30	105.5	4.3	1541	9	US-09-839-996-3 Sequence 3, Appl1
31	105.5	4.3	1541	9	US-10-080-505-3 Sequence 5782, Ap
32	105	4.3	727	10	US-09-815-242-5782 Sequence 12777, A
33	105	4.3	727	10	US-09-815-242-12777 Sequence 5, Appl1
34	105	4.3	1023	9	US-09-884-696-5 Sequence 6888, Ap
35	104	4.2	1344	9	US-09-738-626-6888 Sequence 5797, Ap
36	103.5	4.2	1018	10	US-09-815-242-5797 Sequence 12838, A
37	103.5	4.2	1018	10	US-09-815-242-5797 Sequence 6, Appl1
38	103.5	4.2	2710	9	US-10-011-366-6 Sequence 376, App
39	103	4.2	550	9	US-10-125-692-22 Sequence 11, Appl1
40	102	4.2	999	9	US-09-884-465A-376 Sequence 15, Appl1
41	102	4.2	1391	9	US-10-080-505-15 Sequence 180, App
42	102	4.2	1391	9	US-10-080-505-15 Sequence 12996, A
43	102	4.2	1752	10	US-09-841-132-180 Sequence 5254, Ap
44	102	4.2	6281	10	US-09-815-242-12996 Sequence 5254, Ap
45	101.5	4.1	1027	1	US-08-781-986A-5254

ALIGNMENTS

RESULT 1
US-09-986-682B-3
Sequence 3, Application US/09986682B
Patent No. US20020115182A1
GENERAL INFORMATION:
APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU
KENTKYUO
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/986,682B
FILING DATE: 06-NO. US20020115182A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO. 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 3
US-09-986-682b-3
Query Match 100.0%; Score 2454; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 2.5e-189;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNSGDYKEDYGAHITRADMLKIPGOONSPOKVPQFNASAIKNIDSAKGYKSGNLIDL 60
DB 1 MNSGYKEDYGAHITRADMLKIPGOONSPOKVPQFNASAIKNIDSAKGYKSGNLIDL 60
QY 61 DWSDMPLQONADGTAAHYGHYIVSALAGDPKNSDDTPLHLFYQKVDGTSIDSMKNAGRY 120
DB 61 DWSDMPLQONADGTAAHYGHYIVSALAGDPKNSDDTPLHLFYQKVDGTSIDSMKNAGRY 120
QY 121 FEDMDKFPVNDPYLKYQOEWMSGATLTKDGQVRLFTYDYSNPNEDGGAGANOIISTAQ 180
DB 121 FEDMDKFPVNDPYLKYQOEWMSGATLTKDGQVRLFTYDYSNPNEDGGAGANOIISTAQ 180
QY 181 VMLSDPDAATLKYDGVSDHKSVFDGDDGYVYONIOOFIDEKMWISGDNHTLBDPHYVEDK 240
DB 181 VMLSDPDAATLKYDGVSDHKSVFDGDDGYVYONIOOFIDEKMWISGDNHTLBDPHYVEDK 240
QY 241 GHKYLVEFANTGTDTGCGOSFNNKAYYGGSDVFEQNEKNKLLQSPKKQIASLANGALG 300
DB 241 GHKYLVEFANTGTDTGCGOSFNNKAYYGGSDVFEQNEKNKLLQSPKKQIASLANGALG 300
QY 301 IVELADDTYVKSVMPLVASTNVADEVERANIEFKMNNKWLFTDSGSKMTSDGINDKQV 360
DB 301 IVELADDTYVKSVMPLVASTNVADEVERANIEFKMNNKWLFTDSGSKMTSDGINDKQV 360
QY 361 YMLGPGGSLNGPHNPINETGLVLMNLDPADLTHYSHGCIPIHEGNNVLTSTYNTNG 420
DB 361 YMLGPGGSLNGPHNPINETGLVLMNLDPADLTHYSHGCIPIHEGNNVLTSTYNTNG 420
QY 421 FYPEHHSHRLDKLVNIKSDTSGGENSSGCGQFP 455
DB 421 FYPEHHSHRLDKLVNIKSDTSGGENSSGCGQFP 455
RESULT 2
US-09-995-587A-11
Sequence 11, Application US/09995587A
Patent No. US20020127681A1
GENERAL INFORMATION:
APPLICANT: VAN HIJUM, SACHA ADRIANUS FOKKE TACO
APPLICANT: VAN GEEL-SCHUTTEN, GERITDINA HENDRIKA
APPLICANT: DIJKHUIZEN, LUBBERT
APPLICANT: RAHAOUTI, HAKIM
TITLE OF INVENTION: NOVEL FRUCTOSYLTRANSFERASES
FILE REFERENCE: BO43667-CIP
CURRENT APPLICATION NUMBER: US/09/995,587A
CURRENT FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: 09/604,958
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: EPO 00201872.9
PRIOR FILING DATE: 2000-05-25
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 792
TYPE: PRT
ORGANISM: Lactobacillus reuteri
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (495)..(496)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (737)
OTHER INFORMATION: Thr or Pro
US-09-995-587A-11

Query Match 26.5%; Score 649.5; DB 10; Length 792;
Best Local Similarity 37.8%; Pred. No. 7.2e-44;
Matches 158; Conservative 77; Mismatches 132; Indels 51; Gaps 17;
QY 27 QNSPOFKVPQFNASAIKNIDSAKGYD-KSGNLIDLDVWDSMPLQNA-ADGTAAHYGHYIV 84
DB 203 KODPQYAIPIYSNAKIEKIMNPATYVDAQTKAAHLDVWDSMPLQNPVPTGYVSNWGYDLV 262
QY 85 SALADGPKN-SDDTPLHLFYQKVDGTSIDSMKNAGRVEDMDKFPVNDPYLKYQOEWMSG 143
DB 263 IAMGIIPIPSPTGDNHTIYLLKYGNDPSHMRNAGSIFGTETNV-----FOEWMSG 313
QY 144 SATLTKDGQVRLFTY-----DYSGNPEDGGTGCAGNOIISTAQVNLSDPADATLKYDGYSD 198
DB 314 SAIVVNDGDTIQLEFFISNTSDYKLN-----DQRLATATLNLNVNDNG-VSISKVDN 363
QY 199 HKSVPDGGDYVYONIOOFIDEKMWISGDNHTLBDPHYVE-DKHKYLVEFANTGTDTG 257
DB 364 YQVLEF-GDGFHYQYEQFAN-GKRENDVDCLRDPHYVQLENGDRYLVFEANTGTED-Y 420
QY 258 QGDSQFNNKAYYGGSDVFEQNEKNKLLQSPK-KOIASLANGALGIVELADDTY---VKS 313
DB 421 QSDDOIYMWANGGDPAFIKSSFKLNNKDELALANGALGILKLTNNQSKRYVEV 480
QY 314 MKPLVASNTVADEVERANIEFKMNNKWLFTDSGSKMTSDGINDKQ-----VYMLGPG 366
DB 481 YSPLVSTLMACDEV---XXKLGDKYILEFSYVRVSRGSDRELTDNITVGDNVAMIGYV 536
QY 367 GDSLNGPHNPINETGLVLMNLDPADLTHYSHGCIPI-HEGNNVLTSTYNTNG 421
DB 537 SDSLMGKYYKPLNNSGVLTASVPANMKATYSYVAVPAAGHD--QVLTITSNKRDF 592
RESULT 3
US-09-995-587A-1
Sequence 11, Application US/09995587A
Patent No. US20020127681A1
GENERAL INFORMATION:
APPLICANT: VAN HIJUM, SACHA ADRIANUS FOKKE TACO
APPLICANT: VAN GEEL-SCHUTTEN, GERITDINA HENDRIKA
APPLICANT: DIJKHUIZEN, LUBBERT
APPLICANT: RAHAOUTI, HAKIM
TITLE OF INVENTION: NOVEL FRUCTOSYLTRANSFERASES
FILE REFERENCE: BO43667-CIP
CURRENT APPLICATION NUMBER: US/09/995,587A
CURRENT FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: 09/604,958
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: EPO 00201872.9
PRIOR FILING DATE: 2000-05-25
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 789
TYPE: PRT
ORGANISM: Lactobacillus reuteri
US-09-995-587A-1
Query Match 24.7%; Score 606.5; DB 10; Length 789;
Best Local Similarity 36.2%; Pred. No. 2e-40;
Matches 156; Conservative 75; Mismatches 147; Indels 53; Gaps 17;
QY 31 QKVPQFNASAIKNIDSAKGYDKSGNLID-LDVWDSMPLQNA-DGTAAHYGHYIVSALA 88
DB 229 RYTFEFPKASEKIMNPATTKDAQINTLEPDLVWDSMPLQVQVDRICQVANMNQYOLVIMAM 288
QY 89 GDPKNSDDTPLHLFYQKVDGTSIDSMKNAGRYEEDMDKFPVNDPYLKYQOEWMSGATLT 148
DB 289 GIP-NQNNHNYLLYLNKYGDNELSHMKVGPIFGYNSTAV-----SQEWMSGAVLN 338
QY 149 KDGQVRLFTY---DYSGNPEDGGTGCAGNOIISTAQVNLSDPADATLKYDGVSDHKSVFDGG 206
DB 339 SDNSTQLFYTAVTSDN-----NPNHOKIASATLYLTD-NNGNVSLAQVNRDYVFE-G 390

```

0Y      63 WDSW--PLONADGTAAH-----GHIYSALAGDKRNSDD-----96
      183 WOFMRNDONNGDQVAGATHYLLTAGTNHNOGGAGGWH---SSLGSDYRQAKNYPPIIAGS 239
0Y      97 ----TPLHLYQVKNYDTSIDS-----WKNAGRVE-----DMKRFYNDPYLLKYQTO 1399
Db      240 SGDSGSPMFTYDAEKQKMLINLVLTGNNMAGTENTFOLYKRSFPEILKDLRTSEYSP 299
0Y      140 EMSGSATLTKDG-----QVRLFYTD-----159
Db      300 SGNCAVYITTDKDGSGIYQQTGRSEVRHIGLKDKLPAEGKDYQYQCPNITLPLNN 359
0Y      160 ----YSGNEDG-----GTGA---GNOIIST-----AQNINSOPDAATLKY 193
Db      360 GGNLFEGGOKNKTATVTLSTINOGAGLYPEGNFTYSENNATWOGAGVHHGSDSTYWKV 419
0Y      194 DGVSDHKSVEFDGSGDGTAVYONIOQFIDECKWISGDNHTLDDPHVEKDHGHYLFVE---A 249
Db      420 NGVEENDR-LSKTGKGTILHKAKG-EKKSISVGDGKYLE-OGADQGNKQAFSEIGLVS 476
0Y      250 NTGTDTDYQGDGSFN-NKAYYG-----GSDVEFONEKNKLLOSPKQOIASLANGAL 299
Db      477 GRGTIYQ-LINDKQGFMTDKFYFGFGRGRDLDLNGHSLTF-----KRIQNTDEGAT 523
0Y      300 GIVELADDTYKSVKPLVASTVADVEVERANI EKNNKNWYLFETDSRGSHMTSDGINDKD 3599

```

```

Query Match 128: DB 10; Length 1475;
Best Local Similarity 20.4%; Pred. No. 0.14;
Matches 79; Conservative 48; Mismatches 102; Indels 158; Gaps 22

QY      68  LONADGTANYNG-----YHVSALAGDPKNSDDPTLHLYOKYGDTSIDSKNA 117
      1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:|
Db      113  LKNEGTYA-YYGNDGRRENGCYQMGSCVWKHNNGE-----MSNGLYTID----- 117

QY      118  GHV--FEDMDKFPVNDPRLKYQTQEMSGSATLTGCGVRLFTYDVG----- 163
      1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:|
Db      1177  GGVQVFEDEMG-----YQAK--GKFTVTTADGKIR-YFDKGNGMYRNRFIENEG 1222

QY      164  -----PEDGCTAGNCTTAAVNLISQPAATLLKVDGVDHK--SVFPGGDSVYQNI 215
      1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:|
Db      1223  KMLYLGEDDCAAVTGSOTIN-GQHLXFRANGVQVKGFVTDHGRISYYDGNSSGQLRN-- 127

QY      216  OFI--DEGKWISGDHNTLLRDPHYVEDKGR-----KLYLFKAN----- 250
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db      1280  RFVRNAGQWF-----YFDNNGAVYAGARTINGCOLLYFRANGVQKCEPYTDRT 132

QY      251  --TGTVDGQGDQ-----SPNNKAY-----YGGSDVFQNEKNKLLQSB 287
      1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:|
Db      1329  GRISVYDGNSSGQIRNRFVRNAGQWFFYFDNNGYANTGARTINGCOLLYR----- 137

QY      288  KKQIASLANGALGIVELADDTYVKSYMKPLVASNTVADEVERANIEFKMNNKVLFTTDSG 347

```

```
Db 1379 -----ANG-----VOYKGEFVTRDGRHGRISYYDGNSGDQIRNRFVNNAGOMFYF--DNNG 1426
QY 348 SKMTSDGINDKDYVMGPGDSDLNCPH 374
Db 1427 YAVT-----GARTINGQH 1439

RESULT 6
US-09-815-242-5815
; Sequence 5815, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5815
; LENGTH: 2076
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-5815

Query Match 5.1%; Score 126; DB 10; Length 2076;
Best Local Similarity 23.3%; Pred. No. 0.33;
Matches 111; Conservative 57; Mismatches 204; Indels 104; Gaps 24;

QY 24 PGOONSPOKVPQFNASAIKNIDSAKGYKSGNLDLDVWDSWPLQ--NADGFAANYHGH 82
Db 111 PSAQNNNH---TDGNTTATEIVSNANNNDVVSNNLTALNV---PTKTENGSG---GHL 159
QY 83 IYSALAGPKNSDDPTPLHLFYOKVDTSIDSKNAGRVEDMDKFPVNDPYLKYOTOEWS 142
Db 160 TLKEIOEDVRHSSNNPELV---AIAEPASNRKKRSR-----RAAPADP---NATPADP 207
QY 143 GSATLTKDGQVRLFTYDVS---GNPEDGCTGAGNOIIS-----TAOVNLSQPDAAATL- 191
Db 208 AAAAANGGAPVAITAPYPTPTDPNANNAGNAPNEVLSFDNGIRQSTNRSVPVNVN 267
QY 192 -----KVDSVSDH-----KSVFDSGDSGTIVYQNIQFIDECKWISGDNHTLRDPHY 236
Db 268 NLPGETLLNGKRV-GVFSHAMVRTSMFDSGDKKNYQAQGNVIALGRIHGTDTNDHGFNG 326
QY 237 VED-----KGHKYLFEEANTGTTDGYG-----DQSFNNKAYYGSDVFF-----Q 277
Db 327 IEKALTIVNNSSELIFFENMTTKNGGATNIVLIKNAADTDITAEKTVEGGPTLRLEFKVVD 386
QY 278 NEKNKLOSPPKQIASLANGALGIVELADDTYKVSVMKPLVASNTVADEVERANI--FKM 335
```

```
Db 387 NVRNLIKQIFVPRKND--ITDARGIYOLKKGUYKYSFVDSIGLHSGSHVEERTMDPAT 444
QY 336 NKKWYLFPTDS---RGSKMTSDGINDKDYV--MLGPGDSDLNCPH---PINETGLVANN 387
Db 445 NKRKEFTVITSLKNNNGSASLDOTND-FVYQVQLPEGEYEVNNSLTKRPFPSNNSGDVNV-- 501
QY 388 LDPADLTHYSHCGIHPHPEGNNVVLTSYMTNG-----FYPEHSHLRDLGYN 436
Db 502 -----DMNVTY-----DAANRVITIKSTGGGTANSPARLMPDKILRLRYLRVN 545

RESULT 7
US-09-815-242-12913
; Sequence 12913, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12913
; LENGTH: 2186
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-12913

Query Match 5.1%; Score 126; DB 10; Length 2186;
Best Local Similarity 23.3%; Pred. No. 0.35;
Matches 111; Conservative 57; Mismatches 204; Indels 104; Gaps 24;

QY 24 PGOONSPOKVPQFNASAIKNIDSAKGYKSGNLDLDVWDSWPLQ--NADGFAANYHGH 82
Db 111 PSAQNNNH---TDGNTTATEIVSNANNNDVVSNNLTALNV---PTKTENGSG---GHL 159
QY 83 IYSALAGPKNSDDPTPLHLFYOKVDTSIDSKNAGRVEDMDKFPVNDPYLKYOTOEWS 142
Db 160 TLKEIOEDVRHSSNNPELV---AIAEPASNRKKRSR-----RAAPADP---NATPADP 207
QY 143 GSATLTKDGQVRLFTYDVS---GNPEDGCTGAGNOIIS-----TAOVNLSQPDAAATL- 191
Db 208 AAAAANGGAPVAITAPYPTPTDPNANNAGNAPNEVLSFDNGIRQSTNRSVPVNVN 267
QY 192 -----KVDSVSDH-----KSVFDSGDSGTIVYQNIQFIDECKWISGDNHTLRDPHY 236
Db 268 NLPGETLLNGKRV-GVFSHAMVRTSMFDSGDKKNYQAQGNVIALGRIHGTDTNDHGFNG 326
QY 237 VED-----KGHKYLFEEANTGTTDGYG-----DQSFNNKAYYGSDVFF-----Q 277
```

Db 327 IEKALTVPNSELIFEEFTMTKNGCATNVIKNADETDIAEKTVEGSPTRLRFKVPD 386
QY 278 NEKNLLQSPKQKQSLANGALGIVELADDTYKSVMPKPLVASNVADEVERANI--FKM 335
Db 387 NNRNLIKQFVPRKND--ITDARGIYQKDGKYYYSFVDSIGLHSGSHVEFRTDPAT 444
QY 336 NKKWYLFDS--RGSKMTSDGINDRDYV--MLGPGGDSLNGPHN--PINFETGLVNN 387
Db 445 NKKFTVTSLSKNNNSGASLDFTND--FYQVQLPEGEVEYVNNSLRKDFPSNNSGYDVN-- 501
QY 388 LDPADLTHTYSHCGIPHEGNNVLTSTWTRG-----FYPEHSHLRDLKLVN 436
Db 502 ----DMNVY-----DAANRVTITIKSGGTANSPARLMPDKILDRYKLRVN 545

RESULT 8
US-09-928-457-38
; Sequence 38, Application US/09928457
; Patent No. US20020164603A1
; GENERAL INFORMATION:

APPLICANT:
TITLE OF INVENTION: DNA, specific proteins and peptides
TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method
TITLE OF INVENTION: for obtaining them and their biological application.
NUMBER OF SEQUENCES: 99
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (OEB)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/928,457
FILING DATE: 2001-08-14
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 09/214,759
FILING DATE: 199-12-10
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 1981
TYPE: acide amin, s
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..1981
US-09-928-457-38

Query Match 5.1%; Score 125; DB 9; Length 198;
Best Local Similarity 22.0%; Pred. No. 0.37;
Matches 95; Conservative 62; Mismatches 161; Indels 114; Gaps 23;

QY 68 LONADGTANYHGYHIVS--ALAGDPKNSDTPHLFYQKVDTSIDS--WKNAGRVPE 122
Db 170 IANPNCITVNGGFKFKNVGRGILITIGAPQIGKDGALTGFVROGTLTVGAAGNNDKGA-- 227
QY 123 DMDKFPVNDPYLYKQTOEWSGSATLTKDQOVLFTDYSGNPDGDTGAGNO--IISIA 179
Db 228 DYTGVLAARVALQGLQ--GKNLAVSTGPQV--DVASGEISAGTAGTPTIALDPA 281
QY 180 QVNLQSPDPAATL--KVDVSDHKS-----VFEGGDT-----Y 211
Db 282 ALGCMYADISITLIANEKGVKNAGTLEAKOLIVTSSGRIENSGRIATTAAGTESPT 341
QY 212 QNIQ-----QFDECKWT-----SGDNHTLRDPHYEDKGK-----Y 244
Db 342 LSIETTEKGAAGFTISNGRIESKGLVYETGEDISLRNGAVYQNNGSRPATTVLNAGHN 401
QY 245 LVEFANTGTTD-----GYGDQSFNNKAYYGGSVFQNEKNKLQSPKQIAS----- 293
Db 402 LVIESKTVNNNAKGSANLSAGGRITINDATTIAGSSVYSTGTDELGENTRIIAENVTV 461

QY 294 LANGALG---IVELADDTYKSVMPKPL--VASNTVADEVERANIFKNNKWLFTDSRGSK 349
Db 462 LNSGIGSAVAIEAKDTAHIES--GKPLSLETSTVASNI-----RLNN-----GNIKGSK 509
QY 350 ----MTSDGINDKDYMLGPGGDSLNGPHNPINETGLVNNMLDPAADLTHTYSHC---GI 402
Db 510 QLALLADDNDITAKTT-----NLNTPGNLVYHGGKDLNLNVD--KDLASAASHKSDNA 560
QY 403 PHEGNNVLT 414
Db 561 AHITGTSKTLTA 572

RESULT 9
US-10-066-551-1

; Sequence 1, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Edwards, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the
; TITLE OF INVENTION: prevention and treatment of Neisserial infections
; FILE REFERENCE: 875, 045U1
; CURRENT APPLICATION NUMBER: US/10/066,551
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/344,452
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/310,356
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/266,070
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 1
; LENGTH: 2015
; TYPE: PRP
; ORGANISM: Neisseria gonorrhoeae
US-10-066-551-1

Query Match 5.1%; Score 125; DB 9; Length 2015;
Best Local Similarity 22.0%; Pred. No. 0.38;
Matches 95; Conservative 62; Mismatches 161; Indels 114; Gaps 23;

QY 68 LONADGTANYHGYHIVS--ALAGDPKNSDTPHLFYQKVDTSIDS--WKNAGRVPE 122
Db 170 IANPNCITVNGGFKFKNVGRGILITIGAPQIGKDGALTGFVROGTLTVGAAGNNDKGA-- 227
QY 123 DMDKFPVNDPYLYKQTOEWSGSATLTKDQOVLFTDYSGNPDGDTGAGNO--IISIA 179
Db 228 DYTGVLAARVALQGLQ--GKNLAVSTGPQV--DVASGEISAGTAGTPTIALDPA 281
QY 180 QVNLQSPDPAATL--KVDVSDHKS-----VFEGGDT-----Y 211
Db 282 ALGCMYADISITLIANEKGVKNAGTLEAKOLIVTSSGRIENSGRIATTAAGTESPT 341
QY 212 QNIQ-----QFDECKWT-----SGDNHTLRDPHYEDKGK-----Y 244
Db 342 LSIETTEKGAAGFTISNGRIESKGLVYETGEDISLRNGAVYQNNGSRPATTVLNAGHN 401
QY 245 LVEFANTGTTD-----GYGDQSFNNKAYYGGSVFQNEKNKLQSPKQIAS----- 293
Db 402 LVIESKTVNNNAKGSANLSAGGRITINDATTIAGSSVYSTGTDELGENTRIIAENVTV 461
QY 294 LANGALG---IVELADDTYKSVMPKPL--VASNTVADEVERANIFKNNKWLFTDSRGSK 349
Db 462 LNSGIGSAVAIEAKDTAHIES--GKPLSLETSTVASNI-----RLNN-----GNIKGSK 509
QY 350 ----MTSDGINDKDYMLGPGGDSLNGPHNPINETGLVNNMLDPAADLTHTYSHC---GI 402
Db 510 QLALLADDNDITAKTT-----NLNTPGNLVYHGGKDLNLNVD--KDLASAASHKSDNA 560

OY 403 PHEGNNVLTLS 414
 DB 561 AHITGSKTLTA 572

RESULT 10

US-10-067-385-8
 : Sequence 8, Application US/10067385
 : Patent No. US20020110562A1
 : GENERAL INFORMATION:
 : APPLICANT: Adamou, John
 : TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines
 : FILE REFERENCE: 469201-589
 : CURRENT APPLICATION NUMBER: US/10/067,385
 : CURRENT FILING DATE: 2002-02-05
 : PRIOR APPLICATION NUMBER: US/09/590,991
 : PRIOR FILING DATE: 2000-06-09
 : PRIOR APPLICATION NUMBER: US/60/138,453
 : NUMBER OF SEQ ID NOS: 8
 : SOFTWARE: Patentin Ver. 2.0
 : SEQ ID NO: 8
 : LENGTH: 773
 : TYPE: PRT
 : ORGANISM: Streptococcus pneumoniae
 : US-10-067-385-8

Query Match 4.9%; Score 120; DB 12; Length 773;

Best Local Similarity 22.7%; Pred. No. 0.27; Indels 124; Gaps 25;

Matches 99; Conservative 59; Mismatches 154; Indels 124; Gaps 25;

OY 51 YKSGNLIDLDVWDSWPLQNDGTANYHGY---HVSALAGDPKNSDDTPLHLYOKVG 107
 DB 91 YDDKGIILYD-----DGTDEYETKLEIDEKSKIIYGLVLSKNGHEIL-GRKS 139
 OY 108 DTSIDSWKNA---GRVEDMD-KFVPNDPYLKYOTQEWGSGAT-----LTKDGYRLFY 157
 DB 140 NYS---KNAKYYGNNYKSIKATKYDFHSTKTFEDLYANIINDVGLAFAGDMRLFV 195
 OY 158 TYSGNDPEGCGTGGAGNIISTAOVNLSDPD---AATLKADYSDHKSVDGSDGDTGYXON- 213
 DB 196 KD---NDQ-----KKAEIKIRMEDEKIKETKSETPYVSSGVNLEBGGDSLKK 241
 OY 214 -----IQOFIDEGKMSIDGNHTLDRPHYVEDKHKYLVFEANTGTTDGYOGDOS 262
 DB 242 PDLNFMESGKLYSDSEKQYLLKDNILIR-----KGYALKVTTYNNGKMDLEGN-- 292
 OY 263 FNNKAYYGGSDVFFONEKKKLLQSPKKQJASLANGALGIVELADITYKS-----VVK 315
 DB 293 ---GVYSKEDI-----AKIOKANPNLRLSETTI---YADSHNVEDGRSTQSVLMS 337
 OY 316 PLVASNTVADEVERANIFKMNK-----WTLFTDSRGSKMTSDGINKDYYMLGPGD 368
 DB 338 ALDGNRIITYQY---FTFRMNDKGEALDKGNLVYDS--SKLVLFGRKDKKEY---TGSD 388
 OY 369 SLNGHPNPINETGLVLMNLDPADLTHYSHCGIPHEGNNVLTLSYMNRFGEYDEHHSH 428
 DB 389 KEN--VEAIKEDGSMFLDITKRPVNLMSDKNYP--NPSKSNKI---YVANPEFY----- 434
 OY 429 LMDKLGVMNKSDTSG 444
 DB 435 LRKGI-----SDKGG 444

RESULT 11

US-09-839-996-6
 : Sequence 6, Application US/0983996
 : Publication No. US20030009010A1
 : GENERAL INFORMATION:
 : APPLICANT: St. Geme III, Joseph W.
 : Falkow, Stanley
 : TITLE OF INVENTION: Haemophilus Adherence and Penetration

Protein
 : NUMBER OF SEQUENCES: 9
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 : STREET: 4 Embarcadero Center, Suite 3400
 : CITY: San Francisco
 : STATE: California
 : COUNTRY: United States
 : ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/839,996
 : FILING DATE: 20-Apr-2001
 : CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US/08/296,791
 : FILING DATE: 25-Aug-1994

ATTORNEY/AGENT INFORMATION:
 : NAME: Trecartin, Richard F.
 : REGISTRATION NUMBER: 31,801
 : REFERENCE/DOCKET NUMBER: A-59941/RET/RMS

TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (415) 781-1989
 : TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1848 amino acids

TYPE: amino acid

TOPOLOGY: unknown

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-839-996-6

Query Match 4.7%; Score 116.5; DB 9; Length 1848;

Best Local Similarity 20.8%; Pred. No. 1.6; Indels 155; Gaps 28;

Matches 105; Conservative 64; Mismatches 181; Indels 155; Gaps 28;

OY 25 GOONSPOEKVPOFNAS--AIKNIDSAGYDKSCNLIIDVWDSWPLQNDGTANYHGYH 82
 DB 296 GDSGSPLEFYDREKGMWLFSGYDFWAGYNNKS-----MOENNIYXHEFAEKIYQOYS 348
 OY 83 IYSAL-----AGDPKNSD-----DTPHL----- 100
 DB 349 AGSLIGSNTQYTWQATGSTITGGGEPLSVDLTGKDKRPNHGSITLKGSGTLLNNHI 408
 OY 101 -----LF---YORVGDTSIDSMKNAGRFEDMDKFPV--NDPYLKYOTQEWGSGATL 147
 DB 409 DQAGAGLFEEDGYEVKGTSDSTTWKAGVSVAD-GKTYTWKVNHP--KYDRLAKIGKGTL 465
 OY 148 TRDGYRLFYDTSGNPE-----DG-----GTAGAQIISTAOVNL-SOPDAATLYVD 194
 DB 466 VVEGR-----GKNBGLKVGSDGTLYLKOKADANNKVAQFSQGVISGRSTLVLLMD 516
 OY 195 GVSIDKSYVDG-----GDGTYVONIQOFIDEGKMSIDGNHTLDRPHYVEDKHKYLV 245
 DB 517 KOYDPSNITIFGRGRGLDLNGLSLTFDHIRN-IDGCAHYV--NHNMTNITNITIGES-L 572
 OY 246 VFEANTGTTDGYOGDOS-----FN--NKAYYGGSDVFFONEKKKLLQSPK 288
 DB 573 ITNPMTITSYNLEADDDHPLRIRSIPLYROLYFNODNRSY-----TLKKGASTREEL 625
 OY 289 KOIASLANGALGIVELADITYKSWKPLVASNTVADEVERANIFKMNKMYLFTDSRGS 348
 DB 626 PONSSESENMLYMGRTSDAAKRVNMHI-----NNERNNGF--NGYFGEETKAT 674
 OY 349 K-----MNSDGINDDVYMLGPGDSLNGCPHNPINETGLVLMNLDPADLTHYSHCGI- 402
 DB 675 QNGKLVTFNGKSDONRFL-L-TGGTNLNGDLN-VERGTLFLSGRPT---HARDIAGIS 728

OY 403 -----PH-PEGNNVLTSTYMTNRCF 421
 Db 729 STKDPHFTENNEVVEDDMINRNF 753

RESULT 12
 US-10-080-505-6
 : Sequence 6, Application US/10080505
 : Publication No. US20030073166A1
 : GENERAL INFORMATION:
 : APPLICANT: St. Geme, Joseph W.
 : TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
 : FILE REFERENCE: A-59941-1/RFT/DOF/DHR
 : CURRENT APPLICATION NUMBER: US/10/080, 505
 : PRIOR FILING DATE: 2002-02-22
 : PRIOR APPLICATION NUMBER: US 08/296, 791
 : PRIOR FILING DATE: 1994-10-25
 : PRIOR APPLICATION NUMBER: US 09/839, 996
 : PRIOR FILING DATE: 2001-04-20
 : NUMBER OF SEQ ID NOS: 58
 : SOFTWARE: PatentIn version 3.1
 : SEQ ID NO 6
 : LENGTH: 1848
 : TYPE: PRT
 : ORGANISM: Haemophilus influenzae
 US-10-080-505-6

Query Match 4.7%; Score 116.5; DB 9; Length 1848;
 Best Local Similarity 20.8%; Pred. No. 1.6;
 Matches 105; Conservative 64; Mismatches 181; Indels 155; Gaps 28;

OY 25 GOONSPOFKVPOFNAS--AIKNIDSAGYKSGNLIDLVDWSDPFLQADGTANYHGYH 82
 Db 296 GDSGSPLEFVVDREKGMFLGSDYFMAGYNNKS-----WQENWYKHEAEKYYOYS 348
 OY 83 IVSAL-----AGDPKNSD-----DTPLH----- 100
 Db 349 AGSLIGSTQYTWQATGSTTITGGGEPFLSVLDTGDKPRHGSITLKGSGTLTNHNI 408
 OY 101 -----F-----YQVGDTSIDSKNNGRFEEDMKFVP---NDPYLKYOTQEWGSGATL 147
 Db 409 DQAGAGLEFEGDYEKGTSDSTWKGAGVSAD--GKTVTWKVNHP--KYDLRAKIGKGL 465
 OY 148 TKDQVAFLEYTIDYSGNPE-----DG-----GTGAGNOIISTAVQNL--SOPDAATLKVD 194
 Db 466 VVEKG-----GKNEGGLKAYDGTIVILKQKADANNKQVAFSQVQIVSGRSLVLLND 516
 OY 195 GVSDDHKSVEFG-----GDGTVYONIQOFLDEGKWSGDNHTLDPHYVEDKHKYL 245
 Db 517 KOYDPNSIYEGFRGRLDLNGNSLTFDHIRN--IDDGARVY--NNMTNTSMITTTGES--L 572
 OY 246 VFEANTCTIDGYGDOS-----FN--NKATYGGSDVFFQNEKNKLLOSPK 288
 Db 573 ITNPNITTSYNIEQDDHPLRLRSIPRYOLYFNQDRSY-----TLKGASTRSEL 625
 OY 289 KQIASLANGALGIYELADDTYKSVMKPLVASNTVADEVERANIFKNNKMYLTDSRGS 348
 Db 626 PONSSEENEMWLMGRTSDAKRRVNMHT-----NNERMANGF--NGYFGEERKAT 674
 OY 349 K-----MTSPGINDKDYVMLGPGGDSLNGPHNPINETGLVNLNMLDPADLHTYSHGCI- 402
 Db 675 QNGKLANTFNGKSDQNRFL--TGCTNLNGDLN--VEKGLTFLSGRPTP---HARDIAGIS 728
 OY 403 -----PH-PEGNNVLTSTYMTNRCF 421
 Db 729 STKDPHFTENNEVVEDDMINRNF 753

RESULT 13
 US-09-839-996-4
 : Sequence 4, Application US/09839996
 : Publication No. US2003009010A1
 : GENERAL INFORMATION:

APPLICANT: St. Geme III, Joseph W.
 : Falkow, Stanley
 : TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
 : NUMBER OF SEQUENCES: 9
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 : STREET: 4 Embarcadero Center, Suite 3400
 : CITY: San Francisco
 : STATE: California
 : COUNTRY: United States
 : ZIP: 94111-4187
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/839, 996
 : FILING DATE: 20-Apr-2001
 : CLASSIFICATION: <Unknown>
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US/08/296, 791
 : FILING DATE: 25-AUG-1994
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Treccartin, Richard F.
 : REGISTRATION NUMBER: 31,801
 : REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (415) 781-1989
 : TELEFAX: (415) 398-3249
 : TELE: 910 277299
 : INFORMATION FOR SEQ ID NO: 4:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1545 amino acids
 : TYPE: amino acid
 : TOPOLOGY: unknown
 : SEQUENCE DESCRIPTION: SEQ ID NO: 4:
 US-09-839-996-4

Query Match 4.7%; Score 115.5; DB 9; Length 1545;
 Best Local Similarity 20.8%; Pred. No. 1.6;
 Matches 104; Conservative 55; Mismatches 193; Indels 149; Gaps 23;

OY 25 GOONSPOFKVPOFNAS--AIKNIDSAGYKSGNLIDLVDWSDPFLQADGTANYHGYH 82
 Db 290 GDSGSPLEFVVDREKGMFLGSDYFMAGYNNKS-----WQENWYKHEAEKYYOYS 342
 OY 83 IVSALADGPKNSDPTPLHLYQVGDTSIDSKNNGR-----VFEDMDKFPV 128
 Db 343 AGSL-----IGSKTDYSSWSNGKSTITTGGEKSLVNDLADGKDK-- 381
 OY 129 PNDPYLYQTOEWSGATLTKD---GOVRFYTDYSGNPEDGCT-----GAGNOIIST 178
 Db 382 PN--HGKSVPEEGSGTLTNNDQAGAGLF--FEGDYEVKGTSDNTWKGACVSAVG 436
 OY 179 AQV-----NLSOPDAATLKVDGYSDHKSVEFGDGDGTVYONIQOFLDEGKW----- 223
 Db 437 KTVTWKVNHPQYDLRAKIGKGLTIVEGTGDKKSLKVGCDGTIVILK--DQNGSGOHARASY 495
 OY 224 --ISG--NNHTLRDPHYVE-----GSDVFFQNEKNKLLOSPKQOLA 292
 Db 496 GIVSGRSLVLDKQYDPNSIYEGFRGRLDLNGNSLTFDHIRNIDEGARLVHSHSKH 555
 OY 253 TTIDYQGDQSF--NNKAYY-----GSDVFFQNEKNKLLOSPKQOLA 292
 Db 556 STVITGDNLTDPNNVSIYVKPLEDDNPYAIRQIKYGYQLYF--NEENRYVALKKDAS 614
 OY 293 SLANGALGIYELADDTYKSVMKPLVASNTVADEVERANIFKNNKMYLTDSRGS----- 348
 Db 615 IRSEFPQNRGSSNSWLMYGTAKADQKNA---NNHINERMANGFNGYFGEERKANN 670
 OY 349 -KMTSDGINDKDYVMLGPGGDSLNGPHNPINETGLVNLNMLDPADLHTYSHGCI-P---- 403

```
Db      671 LNVTFKGSSEQNRFL-TGNTNLNGDLN-VQGGTLFLSGRPTP-----HARDIAGISSYRK 724
QY      404 ---HPEGNNVLTSTYMTNRF 421
Db      725 DSHFSENNEVVEEDWIMRNF 745

RESULT 14
US-10-080-505-4
: Sequence 4, Application US/10080505
: Publication No. US20030073166A1
: GENERAL INFORMATION:
: APPLICANT: St. Gene, Joseph W.
: TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
: FILE REFERENCE: A-59941-1/RF/DCF/DHR
: CURRENT APPLICATION NUMBER: US/10/080, 505
: PRIOR FILING DATE: 2002-02-22
: PRIOR APPLICATION NUMBER: US 08/296,791
: PRIOR FILING DATE: 1994-10-25
: PRIOR APPLICATION NUMBER: US 09/839, 996
: PRIOR FILING DATE: 2001-04-20
: NUMBER OF SEQ ID NOS: 58
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 4
: LENGTH: 1545
: TYPE: PRT
: ORGANISM: Haemophilus influenzae
US-10-080-505-4

Query Match      4.7%; Score 115.5; DB 9; Length 1545;
Best Local Similarity 20.8%; Pred. No. 1.6;
Matches 104; Conservative 55; Mismatches 193; Indels 149; Gaps 23;

QY      25 GOQNSPOFKVPOFNAS--AIKNIDSAKGYDKSGLIDLDVMDSWPLQADGTAAHYGH 82
Db      290 GDSSSPLEFYVREKGMFLFSGSYDMAGYMKKS-----WQENNYTPPEAEKITYEYS 342
QY      83 IVSALADPKNSDPTPLHLEFYQKVDTSIDSMKNAGR-----VFEDMDKFV 128
Db      343 AGSL-----IGSKTDYSMSNSNGKSTITGGEKSLNVDLADGKDK-- 381
QY      129 PNDYLYKQTOEWSGATLPKD---GQVRLFTYDYSNPEDGCT-----GAGNQLIST 178
Db      382 PN--HGKSVTFEGSGTLLNNIDOGAGGLE---FEEDYEKGTSDNTTWKAGAVSYAEG 436
QY      179 AQV-----NLSOPDAATLKYDGVSDKSVFDDGDTGYVYONIOOFIDEKRW----- 223
Db      437 KTVWVKVHNPOYDLAKIGKGTLLIVEGIDMKKSLKVGDTIVILK--QOTNSGQHAFASV 495
QY      224 --ISG-DNHTLRDPHYVE-----DKGHKYLVEANTG 252
Db      496 GIVSGRSTLVLDNDKQVDPNSIYFGRGRLDLNGNSLTFDHIINIDEGARLVNHSK 555
QY      253 TTDDYQGDQSF---NNKAYY-----GSDVFPQNEKNKLQSPKQIA 292
Db      556 STVTTTDDNLTDPNNVSIYVYKPLEDDNPYAIQIKYQOLYF--NEENRTYVALKDKAS 614
QY      293 SLANGALGIVELADYTVKSVMKPLVASNTFADEVERANIFKMNKKYVLFDSGS----- 348
Db      615 IRSFEPNREGSNNNSMLYMGTERKADAKNA-----MNHINNRMMGFNGFYGEEGCKNNGN 670
QY      349 -KMTSDGINDKDYVYMLPGDGLNGPHNPINETGLVLMNLDPAADLTHYSHCIP----- 403
Db      671 LNVTFKGSSEQNRFL-TGNTNLNGDLN-VQGGTLFLSGRPTP-----HARDIAGISSYRK 724
QY      404 ---HPEGNNVLTSTYMTNRF 421
Db      725 DSHFSENNEVVEEDWIMRNF 745
```

RESULT 15
US-09-912-020-381

```
: Sequence 381, Application US/09912020
: Patent No. US20020045592A1
: GENERAL INFORMATION:
: APPLICANT: Zyskind, Judith
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Trawick, John
: APPLICANT: Forsyth, R. Allyn
: APPLICANT: Frieolich, Jamie M.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
: FILE REFERENCE: ELITRA.001DVI
: CURRENT APPLICATION NUMBER: US/09/912,020
: CURRENT FILING DATE: 2001-07-23
: PRIOR APPLICATION NUMBER: 09/492,709
: PRIOR FILING DATE: 2000-01-27
: PRIOR APPLICATION NUMBER: 60/117,405
: PRIOR FILING DATE: 1999-01-27
: NUMBER OF SEQ ID NOS: 485
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 381
: LENGTH: 467
: TYPE: PRT
: ORGANISM: E. Coli
US-09-912-020-381

Query Match      4.7%; Score 115; DB 10; Length 467;
Best Local Similarity 20.9%; Pred. No. 0.34;
Matches 101; Conservative 54; Mismatches 135; Indels 194; Gaps 28;
```

```
QY      34 VPOFNASAIK-----NIDSAGKGYDKSGLID--LDVMDSWPLQAD-----GTAAHYG 80
Db      5 VDQEDTEAFTRDKELTSGNIADHNGVSGVPEIHSQDVLNADLVNDRTWDTSKSYNG 64
QY      81 YHIVSALADPKNSDPTPLHLEFYQKVD---TSIDSMKNAGRFEEDMDKFVPNDPLYK 136
Db      65 YGIVA-----MNSDG---HLLTINGNDVNGTELDN-----SSVDVVAATGMYKV 107
QY      137 QTOEWSGATLTL---KDOQVRLFTYDYSN-----PEDGCTGA-----GNQLI----- 176
Db      108 RIDNATGACAIADYKKEI--IYVNDVSNMTFSAANKRADIGATYYQAEQNGNTVVLQOME 166
QY      177 --STAQVNLQOPDAAT-----LKYDGVSDKSVFDDGDTGYVYONIOOFIDE--GKMIS--- 225
Db      167 LTVDAVMALSPSANTMIMNLEQDTV-----GTRLTNSRHGLADNGAMVSYFG 215
QY      226 ---GDNHTLRDPHYVED-----KGKTYLVEANTGTTDGYQGDQ----- 262
Db      216 GNFGDNGTIT---NYDQDVNGIMVGYDTKIDGNNAKWIV--GAAGAFAKGMNDRSGVDQ 271
QY      263 -----FNKKAYYGGSDVF--FQNEKKKLQSPKQIASLANGAL----- 299
Db      272 DSQTAIYSSAHFANNVFVDSGLSYSHFNNDLS-----ATMSNGTYVDGSTNSDA 321
QY      300 -----GIVELADYTVKSVMKPLVASNTFADEVERANIF 333
Db      322 WGRFLKAGYFPKLDGACYVTPYGSVGLFPGSGDDYQLSNMCK-----VDGQSYD 370
QY      334 KMN-----NKWYLETDSRSGSKMT-----SDGINDKDYVYMLPGDGLN--NGPHNPIN 378
Db      371 SMREYELVDAGYFTYSEDAFLPPYFKLAVYVDDSNNDNV-----NGDSIDNGTEGSAV 425
QY      379 ETGL 382
Db      426 RVGL 429
```

Search completed: June 16, 2003, 12:22:38
Job time : 71.9283 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:03:38 ; Search time 55.9426 seconds
(without alignments)
781.893 Million cell updates/sec

Title: US-09-986-682B-3

Perfect score: 2454
Sequence: 1 MNSGDYKEDYGFHHTTRADM.....NIKSDTSGENSSGQGF 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1752	71.4	489	2 F97118	levansucrase [impo
2	1568	63.9	473	2 A25040	levansucrase (EC 2
3	1552	63.2	472	2 J00802	levansucrase (EC 2
4	635	23.9	797	2 B28551	levansucrase (EC 2
5	295	12.0	428	2 H97118	levansucrase [impo
6	277.5	11.3	413	2 JC2520	beta-fructofuranos
7	258	10.5	413	2 S47527	extracellular sucr
8	255	10.4	415	2 S39195	levansucrase - Ery
9	244	9.9	423	2 S33771	levansucrase - Zym
10	243	9.9	423	2 JC2519	levansucrase (EC 2
11	237.5	9.7	430	2 JC7379	levansucrase (EC 2
12	141	5.7	1449	2 B81963	Iga-specific serin
13	139	5.7	1141	2 E89824	hypothetical prote
14	138	5.6	1165	2 T28680	fibrinogen-binding
15	133	5.4	993	2 AE1905	outer membrane sec
16	132.5	5.4	1377	2 T34632	tsh protein - Esch
17	130.5	5.3	2144	2 A97942	metalloprotease
18	129	5.3	728	2 AE0101	probable exported
19	128	5.2	1475	2 B33135	gtfb protein precu
20	127	5.2	2186	2 H89960	hypothetical prote
21	126	5.1	13055	2 T16580	hypothetical prote
22	125	5.1	759	2 AC0368	probable autotrans
23	125	5.1	2015	2 B81989	hypothetical prote
24	121	4.9	555	2 F69312	heterodisulfide re
25	120.5	4.9	3262	2 AH2137	hypothetical prote
26	120	4.9	3240	2 F95074	serine proteinase,
27	119	4.8	363	2 S43159	outer membrane por
28	118	4.8	1797	2 F69195	cell surface glyco
29	117.5	4.8	533	2 C97324	beta-xylosidase, f

30	117.5	4.8	1592	2 A38175	glucosyltransferas
31	116.5	4.7	635	2 S57714	csbp protein - Clo
32	116.5	4.7	2256	2 AD1018	large repetitive p
33	116	4.7	507	2 A40659	elastase (EC 3.4.2
34	116	4.7	582	2 S37047	beta-fructofuranos
35	116	4.7	1457	2 D81019	adhesion and penet
36	115.5	4.7	1545	2 B41859	Iga-specific metal
37	115	4.7	467	2 F64765	Yaitu protein - Esc
38	115	4.7	931	2 S66574	transferrin-bindin
39	115	4.7	1431	2 A45865	dextranucrase (EC
40	115	4.7	1599	2 S22737	glucosyltransferas
41	115	4.7	5291	2 F90696	hypothetical prote
42	114.5	4.7	363	2 AE0616	outer membrane pro
43	114.5	4.7	2020	2 C48399	ABC-type transport
44	114.5	4.7	3110	2 AC0116	probable virulence
45	114	4.6	730	2 F96559	hypothetical prote

ALIGNMENTS

RESULT 1

F97118
levansucrase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: F97118
R:Molling, J., Breton, G., Omelchenko, M.V., Markarova, K.S., Zeng, Q., Gibson, R., L
J., Daly, M.J., Bennett, G.N., Koonin, E.V., Smith, D.R.
J., Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97118
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-489 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79737.1; PID:q15024742; GSPDB:GND0168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1772

Query Match	71.4%; Score 1752; DB 2; Length 489;
Best Local Similarity	71.1%; Pred. No. 2.9e+11;
Matches	323; Conservative 51; Mismatches 78; Indels 2; Gaps 2;
QY	2 NSGDYKEDYGFHHTTRADMILKIPGOONSPOFKYPOFNASAIKIDSAGYDKSGNLIIDD 61
DB	30 NDMNRYKETYGVSHITRYNMKSIPMEONDLKFYKPOFNASTLKNIASAKGYDKGNLIIDD 89
QY	62 VMSWPIQVNAADGTAANYHGVHIVSALAGDPKNSDPTPLHLYOKVGDTSIDSKNMGARY 121
DB	90 VMSWPIQVNAADGTAANYHGVHIVSALAGDPKNSDPTPLHLYOKVGDTSIDSKNMGARY 149
QY	122 EDMDKFVPNDPYLYKYTOEMSGSATLTKDGOVRLFTYDYSGNPEDGGTGANQIISTAOY 181
DB	150 KDSDKYVANDPYLYKYTOEMSGSATLTKDGOVRLFTYDYSGNPEDGGTGANQIISTAOY 209
QY	182 NLSQPAATLKVGVSDHKSVPFGDGTYQNTQQOTIDEGKATISGNNHILRDPHYEDXG 241
DB	210 NLSQPAATLKVGVSDHKSVPFGDGTYQNTQQOTIDEGKATISGNNHILRDPHYEDXG 269
QY	242 HKYLFVEMNTGTGGYOGDQSFNNKRAYYGSDVFPQENRKLK-OSPKQOIASLGAALG 300
DB	270 RKYLFVEMNTGTGGYOGDQSLINKAFYGRSGSFTEKQDQLIDPNKKHDSLAGALG 329
QY	301 IVELADYTVKSVKPLVASNTVADEVERANIERKMNKMYLFTDSGSKWTSIDGINDKDV 360
DB	330 IIEIANDYTLKEMKPLIASNFTVDEIERANVYKMGKRWYLFETDSGSKWTSIDGINDKDV 389
QY	361 YMLGPGGDSINGHNINIEGLVANNLDPADLTHYSHGCIPIHPCGNVNLSTYTNNG 420
DB	390 YMLGFSNSLTGPKPLNKTGLVNLNLPDPTLFTYSHFAVPQTNKGNVNTSYTNNG 449

RESULT 2

A25040
levansucrase (EC 2.4.1.10) sacB precursor - Bacillus subtilis

N.Alternate names: sucrose 6-fructosyl-transferase

C.Species: Bacillus subtilis

C.Date: 16-Aug-1988 #sequence_revision 20-Feb-1995 #text_change 20-Jun-2000

C.Accession: S07309; A25040; I39967; H69702

R.Steinmetz, M.; Le Coq, D.; Aymerich, S.; Genzy-Trebol, G.; Gay, P.

Mol. Gen. Genet. 200, 220-228, 1985

A.Title: The DNA sequence of the gene for the secreted Bacillus subtilis enzyme levansucrase

A.Reference number: S07309; MUID:85295507; PMID:2993818

A.Accession: S07309

A.Molecule type: DNA

A.Residues: 1-473 <STEP>

A.Cross-references: EMBL:X02730; NID:g40118; PIDN:CMA26513.1; PID:g732568

A.Experimental source: Marburg

R.Shimotsu, H.; Henner, D.J.

J.Bacteriol. 168, 380-388, 1986

A>Title: Modulation of bacillus subtilis levansucrase gene expression by sucrose and repressors

A.Reference number: A25040; MUID:87008406; PMID:2428811

A.Accession: A25040

A.Molecule type: DNA

A.Residues: 1-68 <SHI>

A.Cross-references: GB:M14202; NID:g143485; PIDN:AAA22725.1; PID:g143486

B.Fouret, A.; Arnaud, M.; Klier, A.; Rapoport, G.

Biochem. Biophys. Res. Commun. 119, 795-800, 1984

A>Title: Characterization of the precursor form of the exocellular levansucrase from Bacillus subtilis

A.Reference number: I39967; MUID:84178454; PMID:6424671

A.Accession: I39967

A>Status: preliminary; translated from GB/EMBL/DDBJ

A.Molecule type: DNA

A.Residues: 1-11, '1', 13-62 <RES>

A.Cross-references: GB:K01987; NID:g143483; PIDN:AAA22724.1; PID:g143484

R.Kunst, F.; Ogatawara, N.; Moser, T.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Berteaux, C.; Bron, S.; Brönnleut, S.; Bruschel, C.V.; Caldwell, J.; Carter, N.M.; Chazotte, L.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A.Authors: Foulgier, D.; Fritz, C.; Fujita, M.; Furuta, Y.; Funai, S.; Galizzi, A.; Gallier, J.-C.; Harwood, C.R.; Heintz, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hulio, M.F.; Jeon, J.; Konigsberg, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidis, A.; Lardinois, R.; Lauterbach, U.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauelet, A.; Meunier, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, V.; Sato, T.M.; Portellelle, A.; Schlegel, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serotoni, A.; Takanashi, M.; Tamakoshi, A.; Tanaka, T.; Terpieta, P.; Toignon, A.; Tosato, V.; Uchiyama, T.; Winiers, P.; Wipit, A.; Yamamoto, H.; Yanase, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, Y.; Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A.Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A.Reference number: A69380; MUID:96044033; PMID:9384377

A.Accession: H69702

A.Status: nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA

A.Residues: 1-473 <KUN>

A.Cross-references: GB:Z99121; GB:AL009126; NID:g2635827; PIDN:CAB15450.1; PID:g2635958

A.Experimental source: strain 168

C.Genetics:

A.Gene: sacB

C.Keywords: extracellular protein; glycosyltransferase; hexosyltransferase

F.30-473/Product: levansucrase #status predicted <ANT>

Query Match 63.9% Score 1568; DB 2; Length 473;
Best Local Similarity 66.0%; Pred. No. 8,4e-99;
Matches 299; Conservative 53; Mismatches 83; Indels 18; Gaps 5;

OY 2 NSGGYKDYGVGFAHITTRADMILKIPGOQMSPOPKVQGFANASAIKNIDSAKGYDKSGLDLD 61
 | | | | | : | | | | | | | | | | : | | | | | | | | | | | |
DB 33 NKRYKEKYGVSTTTTRDMIOLEPEQOKNEKRVQEPFSSITTKTSKAG-----LD 83

[illegible]

QY	182	NSGQDPAATLKVGVSPDHKSVPFGGDDTVYQNIQGFIDECKWISGDNHTLRDPHYVDKG	241
Db	198	NVSKMSD-DTIAKINVEDHKTIIPD-GDSKTYQNNQGFIDEENYISGDNHTLRDPHYVDKG	255
QY	242	HKYLFEBANTGTDDGCGGGDSFNKKAYVGGSDVFFQNEKKMLQSPKKQIASLANGALGI	301

Db 256 HKYLVFEANTGTENGVOGEESLFFNKAYYGGCTNFEFRKRSQKLOOSAKKRADELANGALGI 315
Oy 302 VELADDTYKSVKMPPLVANSVVADEVERANIFKMNKNKYLFTDSRGSGMTSDGINDXKY 361
Db 316 IELNNDYTLKVKVMPPLTSNTVDEITERANFKMNGKMYLFTDSRGSGMTSDGINDXKY 375
Oy 362 MLGPGGSLNCPHPNPINETGLVLANMLDPAULHTYSHGCIPIHEGNVNVLTYSYKTRNGF 421
Db 376 MLGVNSLSLTPKPYKPLNKTLGLVLOMGIDPNVFTYSHFAVPAQKGNVNVLTYSYKTRNGF 435
Oy 422 YPEHSHLRDKLVNKGSDTSGEENS-SGOGQ 453
Db 436 FEDKATFGPSPFLMNKIKNKTYSVKNSTLEOGQ 468

RESULT 4
B28551
Levansucrase (EC 2.4.1.10) precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-fructosyl transferase
C:Species: Streptococcus mutans
C:Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 15-Oct-1999
C:Accession: B28551
R:Shiroza, T.; Kuramitsu, H.K.
J. Bacteriol. 170, 810-816, 1988
A:Title: Sequence analysis of the Streptococcus mutans fructosyltransferase gene and flanking regions
A:Reference number: A91892; MUID:88115184; PMID:2828325
A:Accession: B28551
A:Molecule type: DNA
A:Residues: 1-797 <SHIR>
A:Cross-references: GB:M18954; NID:G153635; PIDN:AAA8584.1; PID:G153636
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 25.9%; Score 635; DB 2: Length 797;
Best Local Similarity 39.0%; Pred. No. 3.5e-35;
Matches 159; Conservative 74; Mismatches 135; Indels 40; Gaps 19;

Oy 31 QFKVYQNASAKIKNDISAKYD-KSGLNLDLDVDSWPLQNA-DGTANYHGYHIVSLA 88
Db 214 RYALPYFNAKAKIKMKATTRDAQTQGLADLDVDSWVQDAKGEVYNNKGYQLVAMM 273
Oy 89 GDPKNSDDTPLALFYQYKGYDTSIDSMKNGVFEEDMKFVNDPYLKYQTOEMSGSATLT 148
Db 274 GIP-NTNDNHLYLLYNNKGYDNNFDMKNAGSIFG-----YNEPPL--TQEMSGSATVN 323
Oy 149 KDGOVRLFYDYSGNPEDGSGAGNOISTAOVNL--SQPAAATLKVDGVDHKSVEFGG 206
Db 324 EDGSLQLEFYKVDTSKDN---SNNQRLATATVNLGFDODVRLTVE--NDKVLTPRGV 377
Oy 207 DGYVQNIQOFIDEGKMWISGDNHLLRDPHYEDK-GHKYLVFEANTGTGDOGQOSFNN 265
Db 378 MAYHYQSYQGW-RSTFTGADNIAMROHVIYEDENGDKYLVFEASTG-TENYQSGEDQIYN 434
Oy 266 KAYYGGSDVFQNEKNKLLQSPK-QIASLANGALVELADDTYKSV---MKPLVASN 321
Db 435 FTNNGSSAYVVKSLFRLDQODWYNRASMANAAIGILKDGKTFEPVDJFYPLLSST 494
Oy 322 TVAEVERANIFKMNKNKYLFTDSR---GS-----KMTSDSINKDVMMLPGGSLNCP 373
Db 495 MVSSELERPNVAKGDKYLTASRLNHSNNDAWKNKNEVYGD-NVYMLSYVSDQLNG 553
Oy 374 HNPINETGLVLANMLDPAULHTYSHGCIPIHEGNVNVLTYSYKTRNGF 419
Db 554 YKPLNNGSVLTASV-PADWRTATYSYAVVAGSSDILLMTATWTRN 600

RESULT 5
H97118
Levansucrase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: H97118
R:Noelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

..: Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A86900; MUID:21359325; PMID:21359325
A:Accession: H97118
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <KUR>
A:Cross-references: GB:A8601437; PIDN:AAK97939.1; PID:G15024744; GSPDB:GN00168
C:Genetics:
A:Gene: CAC1774

Query Match 12.0%; Score 295; DB 2: Length 428;
Best Local Similarity 27.7%; Pred. No. 1.7e-12;
Matches 112; Conservative 62; Mismatches 147; Indels 84; Gaps 21;

Oy 59 DLDVDSWPLQNAADGTANYHGYHIVSL-----AGDPKNSDDTPLALFYQYKGYDTSIDS 113
Db 66 NLWVMDTPLVKKDGSLAVNGYKVIFFALTSARNVGMKRRHDVAGISYFCSTDG---EN 121
Oy 114 WKNAGRVEDMDKFPNDPYLKYQTOEMSGSATILTKDQVRLFTYDYSGNFEDGCTGAGN 173
Db 122 WYKGLAVNVEDAL-----GSRQWAGSALIDENCMVOFEFYT-----ATGRKG 163
Oy 174 OIISTAOVNLSDP-----DAATLKVDGVDHKSVEFGDGTNYQNIQOFIDEGKMWISGD 227
Db 164 EAVRTFEDRLVKTSTVNDKGVHITCSKHQVILE-PDGYVYOTMOO--AKGPTI--- 217
Oy 228 NHTLDPHYVEDKHK--YLVFEANTGTGDOGQOSFNNKAYYGGSDVFQNEKNKLLQ 285
Db 218 -YSPFDPYFFEDPKTKKDYLFEGNKGCKIEKMKPENIGDKLF-----RKNHIAP 266
Oy 286 SPKRIASLANGALCI-VELADDTYKSVKMPPLVANSVVADEVERANIFKMNKNKYLFTD 344
Db 267 RGVENF-----NGWGIAVAQNKDLTRFKLPLLEAVVNOQLERPQIYMKKNKYLTPI 322
Oy 345 SRGSKMT-SDGINDKDVYMLGPGGSLNCPHPINETGYV--NMULDPAULHTYSHGCI 402
Db 323 SH--KFTYAQGLNGVD-GLYFCGNSLSNKKPLNGNLVITNPNDP---YQTSWYLV 376
Oy 403 PHEGNVNVLTYSYKTRNGFPEHH--SHLR-----DKLGVNKG 439
Db 377 ---SCHDVL-----SFINEYHFNQDLRRGCTFAPLQJSLKG 410

RESULT 6
JC2520
beta-fructofuranosidase (EC 3.2.1.26) - Zymomonas mobilis
N:Alternate names: invertase
C:Species: Zymomonas mobilis
C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 20-Jun-2000
C:Accession: JC2520; PC2377; S50037
R:Kyono, K.; Yanase, H.; Tomomura, K.; Kawasaki, H.; Sakai, T.
Biosci. Biotechnol. Biochem. 59, 289-293, 1995
A:Title: Cloning and characterization of Zymomonas mobilis genes encoding extracellular
A:Reference number: JC2519; MUID:95218269; PMID:7766026
A:Accession: JC2520
A:Molecule type: DNA
A:Residues: 1-413 <KYO>
A:Cross-references: DDBJ:D17524; NID:G809529; PIDN:BA04476.1; PID:G809531
A:Accession: PC2377
A:Molecule type: protein
A:Residues: 1-21 <KY2>
R:Rhee, S.K.
submitted to the EMBL Data Library, March 1994
A:Reference number: S50037
A:Accession: S50037
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-104, 'S', 106-107, 'T', 109-413 <RHE>
A:Cross-references: EMBL:L08094; NID:G468429; PIDN:AAA61488.1; PID:G468430
C:Genetics:

A:Title: Nucleotide sequence of levanucrase gene (levv) of *Zymomonas mobilis* ZM1 (ATCC14028)
A:Reference number: S33771; MUID:93305726; PMID:8318541
A:Accession: S33771
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1423 <SON>
A:Cross-references: GB:L08093; MID:g5680348; PIDN:AAA27695.1; FID:g295699

Query Match	9.98;	Score 244;	DB 2;	Length 423;
Best Local Similarity	24.68;	Pred. No. 4.7e-09;		
Matches 109; Conservative	61;	Mismatches 133;	Indels 140;	Gaps 26

OY	16	TRADILKTRPGOONSQPFQYQENASAIKNI	DSAKYDKSGNILDLD	-----	WVDSW	66
Db	14	TRADAMKY--HTDDPTATMP	-----	HTDYPFWTKXYWMDTW	50	
OY	67	PLONADGTAANYGHYIVALSALGD	-----	PKNSDDPLHLFYQYKVDTSIDS	MKNAGRY	120
Db	51	PLRDRINGOVUSFOGSSVIFALVADRTKYGM	MHNRRNDARIGIFYFSRSG	-----	SNWIFGGHIL	106
OY	121	FEDMDKFPNPNRYLKYQTOEWGSGATLTG	---QYRLTYTDSGNRPEDGTCAGNOIIS	177		
Db	107	LKD-----GANT-----	RSWMSG--CTIMAPGTANSVEFTSVNDP	-----	143	
OY	178	TAQVNLSPDPAATLKVDSGDSKSY-FDS	-----	GDGYVONIQOIFIDEKMNISCD	227	
Db	144	-----SSSVAQOCCGCIYIADDKSWIFD	CFDKYTDLFQADGLTYAD	---YAENFM	150	
OY	228	NHTLDRP--YVEDK--GHKYLVEANTGT	DGYOGDOSFNNAKAYGSGDVFONEKKNLQ	285		
Db	191	--DEFDRPHVFITPKIGKTYALFEGNV	-----	AMERTVAVGEEELGPV	231	
OY	286	SPKKI-----ASLANALALIVE-LADDY	TVKSYMKPLVASNTVADEVETANIFMKNKMYL	341		
Db	232	PRKPTTPRGARICAAIIGIAOLNERTK	EMKLLPLPYTAFGVNDQTERPHVVFQNGILTYL	291		
OY	342	FTDSRGSKMTSDGINDKDYMLGPPG	-----	DSLNGHPNPINTEGLV--NNNLDPADL	353	
Db	292	FTISHHSYV--ADGLS-----	GPQGVYGFVSGNFIQYEPPLNGSGLV	GNFSSDP	340	
OY	394	THYVSHGCIHPREGNNVLTSTM	416			
Db	341	YQAVSHYVA-----TNGLVTSFI	358			

RESULT 10
JC2519
levansucrase (EC 2.4.1.10) precursor - *Zymomonas mobilis*
N:Alternate names: sucrose 6-fructosyltransferase
C:Species: *Zymomonas mobilis*
C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 20-Jun-2000
C:Accession: JC2519, PC2376
R:Kyono, K.; Yamase, H.; Tomomura, K.; Kawasaki, H.; Sakai, T.
Biosci. Biotechnol. Biochem. 59, 289-293, 1995
A:Title: Cloning and characterisation of *Zymomonas mobilis* genes encoding extracellular
A:Reference number: JC2519; MUID:95218269; PMID:7766026
A:Accession: JC2519
A:Molecule type: DNA
A:Residues: 1-423 <KTY>
A:Cross-references: DDBJ: D17524; NID: g809529; PIDN: BAA04475.1; PID: g809530
A:Accession: PC2376
A:Molecule type: protein
A:Residues: 1-19 <KY2>
C:Genetics:
A:Gene: sucE2
C:Keywords: glycosyltransferase; hexosyltransferase
F:1-19/Domain: signal sequence #status predicted <STG>
F:20-432/Product: levansucrase #status predicted <MAT>

Query Match	9.9%	Score 243;	DB 2;	Length 423;
Best Local Similarity	24.4%;	Pred. No. 5.5e-09;		
Matches 108;	Conservative 60;	Mismatches 135;	Indels 140;	Gaps 25

```

QY 16 RADMLKPGQONSQPFKVPFNASAIKNIDSAGYDKSGMLD-----VWDSM 66
   |||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 14 TRADAMKY--HTDDEPTATM-----TIDYDFPIMTDKWTWDTW 50
   |||| : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 PLQNAADGTAANYHGYHIVSALAGD-----PKNSDDPLHLEYOKVGDTSIDSMKNAGRY 120
   || : : : : : : : : : : : : : : : : : : : : : : : : :
Db 51 PLRDINGQVVSFOGVSIFALVYADRTKYGMHNRNDGARIGFYSRGG-----SMNIFGSHL 106
   || : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 FEDMDKFPNDPRLYLYQTOQEMSGSATLTKDG---QVRLPYDYSGPNEDDGTGAGNOIIS 177
   : : : : : : : : : : : : : : : : : : : : : : : : :
Db 107 LKD-----GANP-----KSMWSG--CTTMAPGTANSVEVFFSVNDTP-----143
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 TAOYVLQSPDAILTKVDGVSDBKSY-EDG-----GDCTVQNTQOQFIDEKWTISGD 227
   : : : : : : : : : : : : : : : : : : : : : : : : :
Db 144 -----SSVPAQCKGVIYADCKSVYMGDFPKVTDLQADGLYAAD--YAENFW----190
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 228 NHTLRDPHY--EDKGHLYLVEANTGTTDGTQGDQDSNNKAYYSGDSVFPQNKMKLLQ 285
   |||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 --DFPDHPVFINPEBGKTYALFEENV-----AMORGAVAWEEEGTGV--231
   |||| : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 SPPKOIASLAN---GALGIVE-LADDYTVKSVMKPLVASNTVADEVERANIEFKNNKNWYL 341
   || : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 PRKPTTPRGATYYAAIGIAQALNEAREKMKLLRPLVYAFGVNDQTERPHVVPONGLYTL 291
   || : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 FTDSFGSKMTSDGILNKKDYMLGPGG-----DSLNGPNHPINETGLVL--NMNLDPADL 393
   || : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 FTISHSHSY-ADGLS-----GPDGYVGFSENGIIFGPYPLNGSGIAGLNPSSQP---340
   || : : : : : : : : : : : : : : : : : : : : : : : : :
QY 394 THYSHGCIHPHEGNNVULTSYM 416
   |||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 YQANSHYVA-----TNGLVTSFI 358
   |||| : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 11
JC7379
levansucrase (EC 2.4.1.10) - Acetobacter xylinum, Acetobacter aceti subsp. xylinum
C.Species: Acetobacter xylinum, Acetobacter aceti subsp. xylinum
C.Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 31-Dec-2000
C.Accession: JC7379
R.Tajima, K.; Taniio, T.; Kobayashi, Y.; Kohno, H.; Fujiwara, M.; Shiba, T.; Erita, T.
DNA Res. 7, 237-242, 2000
A.Title: Cloning and sequencing of the levansucrase gene from Acetobacter xylinum NCIT
A.Reference number: JC7379
A.Accession: JC7379
A.Molecule type: DNA
A.Residues: 1-430 <TAJ>
A.Cross-references: DDBJ:AB034152
A.Experimental source: strain NCI 1005
C.Comment: This enzyme, having high polymerase activity, releases glucose and produce
C.Genetics:
A.Gene: lsxA
C.Keywords: glycosyltransferase; hexosyltransferase

Query Match	9.7%	Score 237.5;	DB 2;	Length 430;
Best Local Similarity	23.3%;	Pred. No. 1.3e-08;		
Matches 104; Conservative	59;	Mismatches 139;	Indels 145;	Gaps 22

```

QY 13 AHTRAAMLEKIPGQONSPOCKVPQFNASAIKINDSAKKYDKSGNLIDLDV--MNSWPLON 70
      | | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 16 SHWTLIAAMKVHADDPPTTMYVIDYNP-----YIDQVWQMDGTSGIRA 59
      | | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 71 ADGTAANYHGCHIVSALACDPKNSDDTP-----LHLYQKVDTSIDSKNAGR 119
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 60 ITGTGVAFENDMYMMALVYANRADTGTGTVEGWMHNRNFAIIGYYSRNG--ID-WTIGGR 119
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 120 VEEDMDKFEVNDPLKYOQOEWMSGATLTK--DGQVRLFYTDYSGNPBGDGTGAGNOIIS 177
      : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 116 LLAQSAADLRDP-----EMSGSLVMRAGYKNTFVMEFTSVN----- 150
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 178 TAOVNLISQOPADATLKVDSVHKSVPDQ-----GDGIYYQNIQ--QPTDECKNISC 228
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 151 -TIDINOSVPCSGTSGKJTES-NDEKVEWPEGSKTVEFMSADGVNYVAAAEEDQYFD----- 201
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```

QY 227 DNHTLRDHYVED--KGGKLYLFEANTGTDDXGGDGSFNKK-----AYYG 270
 Db 202 ----FRDHPLPLNADGKTYCLFE--GNPFGVKGKFTLHDAETGAVPPEGYTTPAAGAYG 254
 QY 271 GSDVFEONEKNKLLQSPKQIASLANGALGIVEALADDYTVK-----SYMKPLVASNT 322
 Db 255 AA-----AIGIARLADGAYEKGDPFSRHHWTLPLALYVALG 289
 QY 323 VADEVERANIFKMNKKWYLFDTDSRGSKWTSDDGINDKDYV--MLGPGGSLANGHPNPINETG 381
 Db 290 VNDQTERPHYVFKKNYTYIETFSHSTYTGDSSTPGDGYGVFSYENG--IFGPEPLNLSG 347
 QY 382 LVLMMNLDPADLTH--TSHSGCIRHPEG 407
 Db 348 LVLG---NPSSAPYETYSH--EVDPDG 369

RESULT 12

B81963
 IGA-specific serine endopeptidase (EC 3.4.21.72) NMA0457 [imported] - *Neisseria meningitidis*
 C:Species: *Neisseria meningitidis*
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: B81963
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel,
 Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
 A:Reference number: A81775; MUID:2022556; PMID:10761919
 A:Accession: B81963
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-1449 <PAR>
 A:Cross-references: GB:AL163753; GH:AL157959; NID:g7379120; PIDN:CA83754.1; PID:g737920
 A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: Iga2; NMA0457
 C:Superfamily: Iga-specific metalloendopeptidase
 C:Keywords: hydrolase; serine proteinase

Query Match	5.7%	Score 141;	DB 2;	Length 1449;
Best Local Similarity	22.4%;	Pred. No. 0.27;		
Matches 110; Conservative	68;	Mismatches 193;	Indels 120;	Gaps 27

```

OY      3 GGD--YKEDVGFANHTADMLKIRGQONSPOFKXPQFNA$AKIND$AKGDKRSG----- 55
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      239 SGRBRHANDGPMPIAGA-----AGDSSPMFVYDKTNKWLKGLVLOTQTPYSGRENGF 293
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      56 NLIDLVDWDS-----PLOWAD--GTANVGHYHVSAL-----ACDPKNSD 95
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      294 QLIKRD-----WFYDDIYRGDTHYVFEFPRSRSGHSFTSNNGTGIVETIETNEKVSNP$IKV 349
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      96 DTP$HLEFYQKYDTSID$WKNAGRVFEDMDKFPV-----NDP$YKQTOQEM$S$AL$TK 149
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      350 QT-VRF$ESL$NETDK$EYV$AGV-----NQYRRLNNGENL$P$TDVNGKLLISNNINQ 404
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      150 DGQVRLFYTYDSCNPEDGCGAGNOIISTAQVNL$Q$P$ATLKYDGV$D$K$V$FDG$DGT 209
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      405 GAGGLYEGDEFTVPENNETWOG-----AGVH$ESSTV$W$K$V$ANGV$ANDR-L$KICKGT 457
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      210 VYQNIQGFIDGK$KISGDNHTLRPHYEDK$GHLYVFE-----ANTGTTDGYQDQ$SN- 264
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      458 LHVQAKG-ENQ$SISVSDGTI$LD-QQADDK$KQAF$EIGLV$G$RGTVQ-L$N$DNOFNP 514
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      265 NKAYVY-----GSDVFPQ$NK$K$NLQ$SP$K$O$K$SL$ANGALGIVELAD$DYV$K$VMK 315
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      515 DKLYFGRGRGLDNGHSL$F$-----RIONTDEGA$IV-----NH$N$AT$T$ST 557
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      316 P$V$ANTV$ADEVERANIF$KN-----NKWYLT$D$R$G$K$MT$D$GIN-----DKDYVM 362
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      558 VTIINGNESITQ$PS$K$NINRLNYSKEI$V$ANGF-----G$K$D$T$K$T$G$RLN$V$Q$P$AE$B$RTL 614
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      363 LGPGGDSLNGPHNPIN$ETGLV$LN$NLDPADLIT$HY$HC-----G$P$HEC$GN$NV$LT$S 414
      ||| : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 613 LLSGGTINLNG--NITQTNGKLFESGRPT--PHAYNHLSGWSKMEGIPOGE---LWVDN 666

QY 415 YMTNRGFIPEH 425

Db 667 DWINRTFFKAEN 677

RESULT 13

E898724
Hypothetical protein sdre [imported] - *Staphylococcus aureus* (strain N315)
C:Species: *Staphylococcus aureus*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: E898724
R:Kuroda, M.; Ohte, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mutantli-Il, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsutsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A89758; M0ID:21311952; PMID:11418146
A:Accession: E898724
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1141 <RDB>
A:Cross-references: GB:RA000018; PID:g13700455; PIDN:BA011752.1; GSPDB:GN00149
A:Experimental source: Strain N315
C:Genetics:
A:Gene: sdre

Query Match 5.7%; Score 139; DB 2; Length 1141;
Best Local Similarity 21.7%; Pred. No. 0.26;
Matches 119; Conservative 60; Mismatches 199; Indels 170; Gaps 26

```

QY 3 SGYKEDYGFANITRADM L-----KIPGOANSPOFKVPOFN-----ASAIKNIDS AK 49
Db 464 AGSQVDYDGNIKRLNGSSTIIIDQNTFELIKVYKVPNOQLPOSNRIYDFSGUEDVTSGPONNK 523
QY 50 GYKSGNLDDLDVMD-----SMLQNAADGTAA-----NYHGTH 83
Db 524 SF--SNNVATLFGEDINSAYIIIRVYSKYTPPSDGLDIAQETSMKRTTDXKYGYUYVAYSYN 581
QY 84 VSAIACDDPKKNSDPT--PLHLFYOKVGDPSIDSMKAGHVFEDIMKFEVNPDIYLYKOTQEW 1 414
Db 582 FIYTSNDTGGGDOITYKPEEKL-KIGD-----YWEDVDK-----DOVGOSTDSKEK 629
QY 142 SGS---ATLT-KDGOYRLFYTDYSGNPREDGCTAGCNOIITSAOYNLSOPRA-ATLKVDGY 196
Db 627 PMANVLTLTLPBGDTKRSVTRDANGNHFERGGLKD---ETUYKFEETPAGUYLPPTKVNST 682
QY 197 SDHKSVFDDGDDGVYVNOIQ-----FIDEKMTISGONHILRPHVVEDKGNHLYLFEA 249
Db 683 TDEKDSNGSSITVKNIGKDDMSLDTGFKFEKRYMLGD-----YWEDPTNKDGJODA 734
QY 250 NTGTDTGUYOGDSFNNAKAYVGGSDVFPONEKNKILLOSPPKIOIASLANGALGIVELAD-- 306
Db 735 NEP-----GIKDY-----KYTLKDSITGKILGTTTDDASOKKYFTDLDN 772
QY 307 -DYTVK----SYMKPLVASNTVAD-----VERANIEKMN-----K 338
Db 773 GNTYVEFETPAGYTPPVKNNTTAEDDSNGLTITGVIKDADMMTLDSGFYKTPKYSIADYV 832
QY 339 WYLFETRSKSKMTSDGINDKDYUMLGPGGDSLNGPHNININTEGVLMMNND----- 389
Db 833 WDSNMDKGODSTEEKIKDYKVTLLINKEGEVIG--TTTDCNGYTRREDNLDSCGYKATFE 890
QY 390 -PADLTHYSHC--GIAPREGNNVLT-----SYMTNGBFYREHNSHLRKLGVNINIGSD 441
Db 891 KPAGLGTQYTNTEDEKDDADGGEVDTITDHDHDFLDNGYREE-----D 934
QY 442 TSGGENSS 449
Db 935 TSDSDSDS 942

```


RESULT 14
T28680
fibrinogen-binding protein homolog - *Staphylococcus aureus*
C:Species: *Staphylococcus aureus*
C>Date: 15-Oct-1999 #sequence-revision 15-Oct-1999 #text-change 11-May-2000
C:Accession: T28680
R:Jocsefsson, E.; McCrea, K.; Ni Eidhin, D.; O'Connell, D.; Cox, J.; Hook, M.; Foster, T.
Microbiology 144, 3387-3395, 1998
A:Title: Three new members of the serine-aspartate repeat protein multigene family of *Staphylococcus aureus*
A:Reference number: 220510; MUID:99098700; PMID:9884231
A:Accession: T28680
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1166 <D>
A:Cross-references: EMBL:AJ005647; NID:e1318793; PID:e1318794; PIDN:CA00652.1
C:Genetics:
A:Gene: sdre

Query Match 5.6%; Score 138; DB 2; Length 1166;
Best Local Similarity 22.2%; Pred. No. 0.31;
Matches 122; Conservative 53; Mismatches 201; Indels 174; Gaps 28;

QY 3 SCDYKEYGFRAH-----TRADMLKIPGQNSPQ-----FKVPYNASAIKNDISA 48
DB 469 AGSQVDYGNIKKNGSTIIDONTETIKYKVNSDQOLPQSNRIYDFSOYE-DVTSQPDNK 527
QY 49 KGYDKSGNLJLDLVWD-----SMPLQNDGTAA-----NYGHY 82
DB 528 KSF--SNNVATLDFGDIINSAYIIKVSRYPTSGELDIAOGTSMRTTEKGYNYAGYS 585
QY 83 IVSALACDPKNSDDT--PLHLFYQVGDTSIDSKNAGRVFEDMDKEVPNDPYLYKYOQE 140
DB 586 NFIVTSNDTGGGQGVKPEEKLY-KIGD-----YWEDVDK-----DGVGTDPSKE 630
QY 141 WSGS--ATLT-KDGVRLTYTDSGNPDEGCTAGNOIISTAOVNLSDPA-ATLAVDG 195
DB 631 KPMANVLVTLLTPGDTTKSVRTDANGHYEFGGLKDG---ETVYKFEETPTGYLPTKYN 686
QY 196 VSDHKSVPFGDGDVYONIQ-----FTDEGKMSGDNHTLDRPHYVEDKRGKYLVE 248
DB 687 TTGEEKSNSSSVYKINGDDMSLDGFKPKRYNLGD-----YVMDTKKDGQD 738
QY 249 ANGTGTDYGDOSFNKAYVGGSDVFFQNEKNKLLSPKKQIASLANGALGIVELAD- 306
DB 739 ANEP-----GIKDV-----KVLKDSGTGKVGITTTDASGKYKFTDLD 776
QY 307 --DYTVK-----SVKKPLVANSVTADEVERANI-----FKMNNK----- 338
DB 777 NGNTVVEETPAGTTPV-KNTTADDKDSNGLTTGVIKDADNMTLDRGYKTPKYSIGD 835
QY 339 --WYLFDSRSGSKMTSDGINDKDYMLGPGDSLNGPHNPINETGLVNNMLD----- 389
DB 836 YVWYDSKDKDKOSTENGKIDVYVTLONEKEVIG--TTKTDENGKRYFENLDSGKRYKI 893
QY 390 ---PADLTHYSHC--GIPEGNVNLV-----SYMTNKGYPEHHSHLRDLGAVNKG 439
DB 894 FEKPAGLTQVTVTNTEEDKADGGEVDVTTDHDFTLNDGXFEE----- 938
QY 440 SDTSGGENSS 449
DB 939 -DTSDSDSDS 947

RESULT 15

AE1905
outer membrane secretion protein alr0791 [imported] - *Nostoc* sp. (strain PCC 7120)
C:Species: *Nostoc* sp.
A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence-revision 14-Dec-2001 #text-change 30-Jun-2002
C:Accession: AE1905
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001.

A:Title: Complete genomic sequence of the filamentous nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE1905
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-993 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072748.1; PID:917130136; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0791

Query Match 5.4%; Score 133; DB 2; Length 993;
Best Local Similarity 21.2%; Pred. No. 0.54;
Matches 99; Conservative 56; Mismatches 151; Indels 162; Gaps 21;

QY 54 SGNLIDLDVMDSWPLQANDTAANYHGHVYSALAGDPKNSDDPPLHLYKVD--TSI 111
DB 419 SGNLRDYSAFN-----HSTAAPSPTSLPEFDPYETPNFLF--LIDNNTSA 461
QY 112 DSMKNAGRVFEDMD--KEVPNDPYLYKYOQEMSGSATLTKDGOVRLPYTDSG-NPEDG 167
DB 462 QASSNLQVELQITRTVRREFVFNADY-----NGQANLTFRAMDSNGVAG 505
QY 168 G-TGAGNOIISTAOVNLSDPAATLKYDGVSDHKSVPEDGD--GYVONIQPFIDECKW 223
DB 506 GTTGVAANAVERNATAFSSNTLTASITVSPINNPIQGTGLDKLYGTANEDI----- 556
QY 224 ISGNHTLDRPHYVEDKHKYLVFEANTGTTDGYQGDQ-----SFNNKAYTG--GSDVFFQ 277
DB 557 ING-----NEGNDYLFGRAGNDTLDEGEGNDYLFEGGTGNDTLDEGESSDLYG 604
QY 278 NEKNKLLSPKKQIASLANGALGIVEL--ADYTVKSVMKPLVANSVTADEVER----- 329
DB 605 NEEDNII-----NGGVGNDNLDDGGTDLKRGCTGNDIYTDVTVGDVIEENPNE 653
QY 330 -----ANIFKMNKKWYLFDSRSGSKMTS-----DGIND----- 357
DB 654 GTDKVNSYISWTLGANLENLTILGNTIIDGTGNELDNHIIGNNAVNRLEGSDG-NDWLIG 712
QY 358 --KDYVWYLGPGGSLNRP-----NHPINETGLVYMMNDPRADLTHYSHCGIPH 404
DB 713 KDGNNDLILGGNGNRLNGETGEDPTEGLGNDVYEIISVGVITIEAAP----- 760
QY 405 PEGNNVNLTSYMTNRGEYFEHHSHLRDLGAVNKGSDTSGGENSSGOG 452
DB 761 -AGIDTVISS-----VDWTLGVNLENLTLVGNQATLIG 793

Search completed: June 16, 2003, 12:10:45
Job time : 58.9426 secs

Query Match	Best Local Similarity	Matches	Conservative	Score	DB 1:	Length	DB 2:	Score	DB 3:	Length	DB 4:	Score	DB 5:	Length	DB 6:	Score	DB 7:	Length	DB 8:	Score	DB 9:	Length	DB 10:	Score	DB 11:	Length	DB 12:	Score	DB 13:	Length	DB 14:	Score	DB 15:	Length	DB 16:	Score	DB 17:	Length	DB 18:	Score	DB 19:	Length	DB 20:	Score	DB 21:	Length	DB 22:	Score	DB 23:	Length	DB 24:	Score	DB 25:	Length	DB 26:	Score	DB 27:	Length	DB 28:	Score	DB 29:	Length	DB 30:	Score	DB 31:	Length	DB 32:	Score	DB 33:	Length	DB 34:	Score	DB 35:	Length	DB 36:	Score	DB 37:	Length	DB 38:	Score	DB 39:	Length	DB 40:	Score	DB 41:	Length	DB 42:	Score	DB 43:	Length	DB 44:	Score	DB 45:	Length	DB 46:	Score	DB 47:	Length	DB 48:	Score	DB 49:	Length	DB 50:	Score	DB 51:	Length	DB 52:	Score	DB 53:	Length	DB 54:	Score	DB 55:	Length	DB 56:	Score	DB 57:	Length	DB 58:	Score	DB 59:	Length	DB 60:	Score	DB 61:	Length	DB 62:	Score	DB 63:	Length	DB 64:	Score	DB 65:	Length	DB 66:	Score	DB 67:	Length	DB 68:	Score	DB 69:	Length	DB 70:	Score	DB 71:	Length	DB 72:	Score	DB 73:	Length	DB 74:	Score	DB 75:	Length	DB 76:	Score	DB 77:	Length	DB 78:	Score	DB 79:	Length	DB 80:	Score	DB 81:	Length	DB 82:	Score	DB 83:	Length	DB 84:	Score	DB 85:	Length	DB 86:	Score	DB 87:	Length	DB 88:	Score	DB 89:	Length	DB 90:	Score	DB 91:	Length	DB 92:	Score	DB 93:	Length	DB 94:	Score	DB 95:	Length	DB 96:	Score	DB 97:	Length	DB 98:	Score	DB 99:	Length	DB 100:	Score	DB 101:	Length	DB 102:	Score	DB 103:	Length	DB 104:	Score	DB 105:	Length	DB 106:	Score	DB 107:	Length	DB 108:	Score	DB 109:	Length	DB 110:	Score	DB 111:	Length	DB 112:	Score	DB 113:	Length	DB 114:	Score	DB 115:	Length	DB 116:	Score	DB 117:	Length	DB 118:	Score	DB 119:	Length	DB 120:	Score	DB 121:	Length	DB 122:	Score	DB 123:	Length	DB 124:	Score	DB 125:	Length	DB 126:	Score	DB 127:	Length	DB 128:	Score	DB 129:	Length	DB 130:	Score	DB 131:	Length	DB 132:	Score	DB 133:	Length	DB 134:	Score	DB 135:	Length	DB 136:	Score	DB 137:	Length	DB 138:	Score	DB 139:	Length	DB 140:	Score	DB 141:	Length	DB 142:	Score	DB 143:	Length	DB 144:	Score	DB 145:	Length	DB 146:	Score	DB 147:	Length	DB 148:	Score	DB 149:	Length	DB 150:	Score	DB 151:	Length	DB 152:	Score	DB 153:	Length	DB 154:	Score	DB 155:	Length	DB 156:	Score	DB 157:	Length	DB 158:	Score	DB 159:	Length	DB 160:	Score	DB 161:	Length	DB 162:	Score	DB 163:	Length	DB 164:	Score	DB 165:	Length	DB 166:	Score	DB 167:	Length	DB 168:	Score	DB 169:	Length	DB 170:	Score	DB 171:	Length	DB 172:	Score	DB 173:	Length	DB 174:	Score	DB 175:	Length	DB 176:	Score	DB 177:	Length	DB 178:	Score	DB 179:	Length	DB 180:	Score	DB 181:	Length	DB 182:	Score	DB 183:	Length	DB 184:	Score	DB 185:	Length	DB 186:	Score	DB 187:	Length	DB 188:	Score	DB 189:	Length	DB 190:	Score	DB 191:	Length	DB 192:	Score	DB 193:	Length	DB 194:	Score	DB 195:	Length	DB 196:	Score	DB 197:	Length	DB 198:	Score	DB 199:	Length	DB 200:	Score	DB 201:	Length	DB 202:	Score	DB 203:	Length	DB 204:	Score	DB 205:	Length	DB 206:	Score	DB 207:	Length	DB 208:	Score	DB 209:	Length	DB 210:	Score	DB 211:	Length	DB 212:	Score	DB 213:	Length	DB 214:	Score	DB 215:	Length	DB 216:	Score	DB 217:	Length	DB 218:	Score	DB 219:	Length	DB 220:	Score	DB 221:	Length	DB 222:	Score	DB 223:	Length	DB 224:	Score	DB 225:	Length	DB 226:	Score	DB 227:	Length	DB 228:	Score	DB 229:	Length	DB 230:	Score	DB 231:	Length	DB 232:	Score	DB 233:	Length	DB 234:	Score	DB 235:	Length	DB 236:	Score	DB 237:	Length	DB 238:	Score	DB 239:	Length	DB 240:	Score	DB 241:	Length	DB 242:	Score	DB 243:	Length	DB 244:	Score	DB 245:	Length	DB 246:	Score	DB 247:	Length	DB 248:	Score	DB 249:	Length	DB 250:	Score	DB 251:	Length	DB 252:	Score	DB 253:	Length	DB 254:	Score	DB 255:	Length	DB 256:	Score	DB 257:	Length	DB 258:	Score	DB 259:	Length	DB 260:	Score
-------------	-----------------------	---------	--------------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------

```

0Y 302 VELADDYVSKVMKPIVASNTVADEVANIEFKNNKNNKMYLFTDSRSGSKMTSDGINDKDYV 361
Db 316 IELMDYLLKRYKMKPPLIASNSVTVDEIEEAVNFKMNGWYLETFTDSRSGSKMTIDGITSNDIY 375
0Y 362 MLGFGGSDLNCPHPININTEGTGYLVLMNMLDPADLTFTYSHGCIPEHNGNVLTYSTMYTNGCF 421
Db 376 MLGVYNSNLTPGYRPLNKTGTGLYKMDLDPNDVFTYTHSHFANVPQAKGNVYITSTMYTNGCF 435
0Y 422 YPHHSHLDRKLGVINKSGTSGGENSE-SCGGQ 453
Db 436 YADKOSTEAPSPFLINIKGKTKSVKSDILEGGQ 468

RESULT 2
SACB_BACAM STANDARD: PRT: 472 AA.
AC P21130:
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levensucrase precursor (EC 2.4.1.10) (beta-D-fructofuranosyl
GN transferase) (Sucrose 6-fructosyl transferase).
OS Bacillus amyloliquefaciens.
OC Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23844;
RA MEDLINE=91092506; PubMed=2265762;
RT Tang L.B., Lenstra R., Borchert T.V., Vasantha N.:
RT "Isolation and characterization of levansucrase-encoding gene from
RT Bacillus amyloliquefaciens.";
RL Gene 96:89-93(1990).
CC -i- CATALYTIC ACTIVITY: Sucrose + {(2,6)-beta-D-fructosyl}(N) =
CC glucose + {(2,6)-beta-D-fructosyl}(N+1).
CC -i- SUBCELLULAR LOCATION: Secreted.
CC -i- INDUCTION: BY SUCROSE.
CC -i- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52988; CAA37179.1; -
DR PIR: J00802; J00802.
DR PIR: S11739; S11739.
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
DR TRANSFERASE: Glycosyltransferase; Signal.
FT SIGNAL 1
FT CHAIN 29
FT LEVANSUCRASE.
SQ SEQUENCE 472 AA: 52859 MW: F38592D272677E7D CRC64:

Query Match 63.2%; Score 1552; DB 1; Length 472;
Best Local Similarity 65.3%; Pred. No. 1,4e-92;
Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;

0Y 2 NSGQYKEYGCAHITRADMELKIPQOQNSPQKRYVQFNASAIKNIKDSKAGKVDKSNLIDLD 61
Db 33 NQAKYKEYGSHTRTHDMIDIPRQOQNEKYVQFQDSTLKNIKESAKG-----LD 83
0Y 62 VWDMSPLQNAAGCTAANYHGYIVSALGADPKNSDDPFLHLEYOKVGGTSTDSMKNAGRVF 121
Db 84 VWDMSPLQNAAGCTAANYHGYIVSALGADPKNSDDPSTLYMYQVGGNDSIDSMKNAGRVF 143
0Y 122 EMDKRFVNDPYLVLEYQTOEWSGSAATLTRKDGVRLEFYDYDSGPNEDGCTGAGNQIITRAV 181

```

```

Db 144 KDSKDFANDPILKDOJEMSGSATFTSDGKIRLFYTDYSGK-----HYGKOSLTTAcy 197
Oy 182 NLSOPDAATLKVDYGSVDFDGGGTYONIOOFIDEGKMLSGDNHLLRDPHYVEDG 241
Db 198 NWSKSD-DTLKINGVEDHKTIFD-GDGKTYONVOQFIDEGANTSGDNHLLRDPHYVEDG 255
Oy 242 HKLYVEANTGTDDYOGDQSFNNKAYYGGSDVFEFONEKNKLLQSPKKOIASLANGALGI 301
Db 256 HKLYVEANTGTENGVOGESLFFNKAYYGGGINTFFRKESQKLQOSAKKDAELANGALGI 315
Oy 302 VELADDTYKSVKPLVASNTVADEVERANIEFKMKNKWLFTFDSRGSKMTSGCINDKDY 361
Db 316 IELNDYTLKKYWKPLIASNTVTDEIERANVEFMKNKWLFTFDSRGSKMTIDGINSNDY 375
Oy 362 MUGPGDLSNGPHNPINETGVLVNMNLDPADLTHYSHGCIPIHPEGNNVLTSMYNRGF 421
Db 376 MGIYVNSLTPGPKPLNKTGLVLMKDLDPNDVTFYSHFAVPOAKGNNAVITSYMTNRGF 435
Oy 422 YPEHSHLRDLKGLVNIKSGDTSIGENS-SGQGO 453
Db 436 FEDKATFGPSPFLMNKMTSVKNSILEQGO 468

```

RESULT 3

```

SACB_BACST STANDARD: PRT: 473 AA.
ID SACB_BACST STANDARD: PRT: 473 AA.
AC P94468;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levensucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
transiferase) (Sucrose 6-fructosyl transferase).
DE SABC OR SURB.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12980;
RX MEDLINE=98007870; PubMed=9349714;
RA Li Y., Triccas J.A., Ferencik T.;
RT "A novel levansucrase-levanase gene cluster in Bacillus
steatothermophilus ATCC12980."
RL Biochim. Biophys. Acta 1353:203-208(1997).
CC -1- CATALYTIC ACTIVITY: Sucrose + {(2,6)-beta-D-fructosyl}(N) =
glucose + {(2,6)-beta-D-fructosyl}(N+1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U34874; AAB97111.1; -
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
KW Transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 29 BY SIMILARITY.
FT CHAIN 30 473 LEVANSUCRASE.
SQ SEQUENCE 473 AA: 52854 MW: 1257A272EC2B771E CRC64;

```

```

Query Match 62.2%; Score 1526.5; DB 1; Length 473;
Best Local Similarity 65.2%; Pred. No. 5.9e-91;
Matches 288; Conservative 52; Mismatches 85; Indels 17; Gaps 4;

```

```

Oy 2 NSGDKEYGFAHTRADMLKIPGOONSPOKVPQFNMSAIKINDSAGXGDSGNLTDID 61
Db 33 NOKPKETTYGISHITRHMDLQIPBOOKNEKTYOVPEPDSSTIKNISAGK-----LD 83

```

```

Oy 62 VWSWPLQADGTAAANYHGYHIVSALAGDPKNSDDTPLLFFYOKYGDTSIDSMKNAGRYE 121
Db 84 VWSWPLQADGTAAANYHGYHIVSALAGDPKNSDDTPLLFFYOKYGDTSIDSMKNAGRYE 143
Oy 122 EDMKRFVNDPPLKTYOTQEMSGSATFTSDGKIRLFYTDYSGK-----HYGKOTLTTAcy 181
Db 144 KDSKDFANDPILKDOJEMSGSATFTSDGKIRLFYTDYSGK-----HYGKOTLTTAcy 197
Oy 182 NLSOPDAATLKVDYGSVDFDGGGTYONIOOFIDEGKMLSGDNHLLRDPHYVEDG 241
Db 198 NWSKSD-DTLKINGVEDHKTIFD-GDGKTYONVOQFIDEGANTSGDNHLLRDPHYVEDG 255
Oy 242 HKLYVEANTGTDDYOGDQSFNNKAYYGGSDVFEFONEKNKLLQSPKKOIASLANGALGI 301
Db 256 HKLYVEANTGTENGVOGESLFFNKAYYGGGINTFFRKESQKLQOSAKKDAELANGALGI 315
Oy 302 VELADDTYKSVKPLVASNTVADEVERANIEFKMKNKWLFTFDSRGSKMTSGCINDKDY 361
Db 316 IELNDYTLKKYWKPLIASNTVTDEIERANVEFMKNKWLFTFDSRGSQMTIDGITSNDY 375
Oy 362 MUGPGDLSNGPHNPINETGVLVNMNLDPADLTHYSHGCIPIHPEGNNVLTSMYNRGF 421
Db 376 MGIYVNSLTPGPKPLNKTGLVLMKDLDPNDVTFYSHFAVPOAKGNNAVITSYMTNRGF 435
Oy 422 YPEHSHLRDLKGLVNIKSGDTS 443
Db 436 YADKOSTFAPSPFLMNKMTSVKNSILEQGO 457

```

RESULT 4

```

SACB_STRSL STANDARD: PRT: 969 AA.
ID SACB_STRSL STANDARD: PRT: 969 AA.
AC 055242;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levensucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
transiferase) (Sucrose 6-fructosyl transferase).
DE FTF.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25975;
RX MEDLINE=93322352; PubMed=8331080;
RA Rathnam C., Giffard P.M., Jacques N.A.;
RT "The cell-bound fructosyltransferase of Streptococcus salivarius: the
carboxyl terminus specifies attachment in a Streptococcus gordonii
model system."
RL J. Bacteriol. 175:4520-4527(1993).
CC -1- CATALYTIC ACTIVITY: Sucrose + {(2,6)-beta-D-fructosyl}(N) =
glucose + {(2,6)-beta-D-fructosyl}(N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED. CELL-WALL BOUND.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: I08445; AAA71925.1; -
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
DR TIGRPFAMs: TIGR01168; YSTRK_signal; 1.
KW Transferase; Glycosyltransferase; Signal; Cell wall.
FT SIGNAL 1 969 POTENTIAL.
FT CHAIN 2 969 LEVANSUCRASE.
SQ SEQUENCE 969 AA: 103983 MW: D389B5B32ACF735A CRC64;

```

Query Match 25.9%; Score 636; DB 1; Length 969;
 Best Local Similarity 36.7%; Pred. No. 2e-33;
 Matches 168; Conservative 78; Mismatches 128; Indels 84; Gaps 23;

26 QONSQFQKPPQNASIKKIDSAKGYD-KSGNLIDLDVDSWPLQNA-DGTAAHYHGYH 83
 249 KODS-KLAVYFKADITIMMPAFNTVDAQTKEEDLDVDSWPLQNAKSGVSNMNGYQL 307
 84 VALAGDKPKNSDDTPLHFLYOKVGDPTSIDSMKNAGRVF-----EDMDKRVPPNDPLYKYOT 138
 308 VISMAGAP-NKSNHIIYLLRYKYGNDFTHKMKNAGPIFGYNALED-----D 352
 139 QWSSGATITKDGQVRLFTDYSGNPEDGSGAGNOIISTAOVNLSDPAATLKVDGVS 198
 353 QOMSGSATVNSDGSIOLYTK---NDTSGGKLNMOOLAS-ATLNLAVENDEVVIRKSVEND 408
 199 HKSVPDGGGTYQYQNOQFID-----EGKMSIDNITLDRPHVDEKGRHYLFEEAN 250
 409 H-ILFSGYHYQSYPRKMSFTDDHNDHNDPDRDNTCLDRPHILENGSRYLFEESN 466
 251 TGTGQYQDQSGFNKAYYSGSVF---FQY-ENKLLQSPKQIATSLANGALGIVELA 305
 467 TG-DEYQGEKQITKMSNYGGDPAFLKSFLLNVNKKLYN---LASMANGSIGILKLD 521
 306 DDY---TVKSYMKPLVASFVADEVERANIFKMNKNWLYFTDSRSGKMTS-----DG 354
 522 DNEKNPSVALTYPTLVTSHMTVDEVERPQVYKMGKGYLFTASIRKSTDAEGTVARBA 581
 355 INDKVYMKPGGDSLNQHNPINTEGLVLMNLDPAD-LTHYSHCGIPHEGNN--VV 411
 582 VSD-DVVMGEPVSDSRGKRYPLNSGVVLTASV-PADWRTSTSYSAVP-VEGSSDTLL 638
 412 LFSYMTNRGFYEHSHLRLDKLVNKGSDTSGGENSS 449
 639 VTSYMTNRG-----GI-----AGAENKS 656

Db

RESULT 5
 SADB_STRMU STANDARD; PRT; 797 AA.

AC P11701;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Levansucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
 transferase) (Sucrose 6-fructosyl transferase).
 GN FTF.
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=GS-5;
 RA MEDLINE=88115184; PubMed=2828325;
 RA Shitoto T., Kuramitsu H.K.;
 RT "Sequence analysis of the Streptococcus mutans fructosyltransferase
 gene and flanking regions.";
 RL J. Bacteriol. 170:810-816(1988).
 CC -i- CATALYTIC ACTIVITY: Sucrose + ((2,6)-beta-D-fructosyl)(N) =
 glucose + ((2,6)-beta-D-fructosyl)(N+1).
 CC -i- SUBCELLULAR LOCATION: Secreted.
 CC -i- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL: M18954; AAA88584.1; -
 DR PIR: B28551; B28551.
 DR InterPro: IPR003469; Glyco_hydro.68.
 DR Pfam: PF02435; Glyco_hydro.68; 1.
 KW Transferase; Glycosyltransferase; Signal.
 FT SIGNAL 1 34
 FT CHAIN 35 797 LEVANSUCRASE.
 SQ SEQUENCE 797 AA; 87574 MW; 6DDE057465DB4042 CRC64;

Query Match 25.9%; Score 635; DB 1; Length 797;
 Best Local Similarity 39.0%; Pred. No. 1.8e-33;
 Matches 159; Conservative 74; Mismatches 135; Indels 40; Gaps 19;

31 QKRVQFNASAIKKNIDSAKGYD-KSGNLIDLDVDSWPLQNA-DGTAAHYHGYH 88
 214 RYAIYFNKAKIKKNAKATTPAQIQQLADLDVDSWPLQNAKTEVINMNGYQLVAMM 273
 89 GPKNSDDTPLHFLYOKVGDPTSIDSMKNAGRVFEDMDKRVPPNDPLYKYOTQOMSGSATLT 148
 274 GIP-MTNDHIIYLLNKNYKGDNNFDHKNKAGSIFG-----YNEPL---TQMSGSATVN 323
 149 KDGQVRLFTDYSGNPEDGSGAGNOIISTAOVNL-SQPAATLKVDGVSDBKSVFDDG 206
 324 EDGSLQLEFTTKYDTSQDN---SNNQRLATATVNLGFDQDQVRLISVE--NDKVLTPGCV 377
 207 DGTVYQNIQFIDEKMWISGDNHNLDRPHYEDK-GHKRYLFEANTGTTDGYQDQSFNN 265
 378 MAYHYQSQQW--RSTFGADNIARDPHVIDEDNGDRYLVEASTG--TENYQGEDQIYN 434
 266 KAYYGSVDFPQNEKNKLLQSPK-QIASLANGALGIVELADYVKS---MKPLVASN 321
 435 FTVYQSSASVYKSLFRFLDDQDMYNRASMANAIGILKLDKDKTPPEVDQFYTPPLSST 494
 322 TVADEVERANIFKMNKNWLYFTDSR---GS-----KMSDQINDKDYMLGPGGDSLNGP 373
 495 WSDLEERNVYKLDQKYLLFTASRLNHGNSNDANKNAKEVYGD-NVMVLGYSQQLTNG 553
 374 HNPINTEGLVLMNLDPAD-LTHYSHCGIP-HPEGNNVLTSTYMTNR 419
 554 YKPLNNSGVVLTASV-PADWRTATYSYVAVPAGSSDTLLMAYMTNR 600

Db

RESULT 6
 SADB_RAHAO STANDARD; PRT; 415 AA.

AC 054435;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Levansucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase).
 GN SADB.
 OS Rahnella aquatilis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Rahnella.
 OX NCBI_TaxID=34038;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC 33071;
 RA MEDLINE=99127094; PubMed=9928133;
 RA Song K.B., Seo J.W., Kim M.G., Rhee S.K.;
 RT "Levansucrase of Rahnella aquatilis ATCC33071. Gene cloning,
 expression, and levans formation.";
 RL Ann. N.Y. Acad. Sci. 864:506-511(1998).
 CC -i- CATALYTIC ACTIVITY: Sucrose + ((2,6)-beta-D-fructosyl)(N) =
 glucose + ((2,6)-beta-D-fructosyl)(N+1).
 CC -i- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: U91484: AAC36458.1: -
 DR InterPro: IPR003469: Glyco_hydro_68.
 DR Pfam: PF02435: Glyco_hydro_68; 1.
 DR Transferrase: Glycosyltransferase.
 KW
 SQ SEQUENCE 415 AA: 45939 MW: E028828813D13A74 CRC64;

Query Match 12.2%; Score 299; DB 1; Length 415;
 Best Local Similarity 26.2%; Pred. No. 2.8e-12;
 Matches 116; Conservative 60; Mismatches 141; Indels 126; Gaps 24;

16 TRADMLKIPGQNSPQFVKVPOFNASAIKINDSAKDYKSGNLIJDD-----VMSWM 66
 12 TRADALKV--NENDPTTQP-----IVDAPFVNSDEVITWDTM 48
 67 PLQANADGTAAVHGHIYSALAGDPKNSDDTPLHLYOKVGDTSIDS-WKN----- 116
 49 PLRSIDGTIVSVDSGWSVFTLTAQRNNNNSE---YLDACGNVDITSDMNNRHRARICY 104
 117 -----ACRFEDMDKFPVNDPLKQTOEMSSS-ATLTKDGOVRLFTYDSGNP 164
 105 WYSRTGKDWIFGGRVMAE-----GVSPTSREMACTPILLNEDGDIDLYTCVT--- 152
 165 EDGCGAGNQIISTPAQVNLSDPAATLKVDSVSDHKSVDGDDGTVYONIDQFIDEKMI 224
 153 -----PGATIAKVRGKVLTSSEGVTLA--GFNEKSLF-SADGYVYOTESO---NPTYN 200
 225 SGNHTLDRPHYVEDKGRKYLVEFANTGTCTDGYOGDOSFNKKAYVYGSDFEQNEKNKL 284
 201 FRDPSPTFDPP---DGKLYMVEGNVA-----GER-----GSHVIGKQMGTL- 240
 285 QSPKQOIASLNGALGIYELA-----DOYTVKSVMKPLVASNTVADEVERANIFKANN 337
 241 -PGRHVGNAKRYOAGCIGMAVANDLSGDEW---ELPLPLTAVGVNDQTERPHRFVODG 296
 338 KMYLTFDSRSGKMT-SDGINDKD-VYMLGPGDLSLNGPHNPINETGLVNNHLDPAJLTH 395
 297 KYVLETFISH-KFTYADGLTGRPDGY--GLSLDNLGTGYSPMNSGLVVLGN--PPSQPQ 350
 396 TYSHCGIPHEGNNVLTSTYMTN 418
 351 TYSHCVMP-----NGLVTSFIDN 368

RESULT 7

SACB_ACEDI STANDARD: PRT; 584 AA.

AC 043998;
 DT 15-DEC-1998 (Rel. 37; Created)
 DT 15-DEC-1998 (Rel. 37; Last sequence update)
 DT 15-JUN-2002 (Rel. 41; Last annotation update)
 DE Levanosucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
 transferase) (Sucrose 6-fructosyl transferase).
 GN LSDA.
 OS Acetobacter diazotrophicus (Gluconacetobacter diazotrophicus).
 CC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
 CC Gluconacetobacter.
 CC NCBI_TaxID=33996;
 RN (1)
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 52-61.
 RC STRAIN-SRT4 / CBS 550.94;
 RX MEDLINE=96253999; PubMed=8704949;
 RA Arieta J., Hernandez L., Coego A., Suarez V., Balmori E.,
 Mendez C., Petit-Glatron M.-F., Chambert R., Selman-Housein G.;
 RT "Molecular characterization of the levanosucrase gene from the
 endophytic sugarcane bacterium Acetobacter diazotrophicus SRT4.";
 RL Microbiology 142:1077-1085(1996).
 CC -1- FUNCTION: RELEASES FRUCTOOLIGOSACCHARIDES AND LEVAN, A HIGH-
 MOLECULAR-MASS FRUCTOSYL POLYMER, FROM SUCROSE. IT ACTS MORE AS A
 SUCROSE HYDROLASE THAN AS A FRUCTAN POLYMERASE.
 CC -1- CATALYTIC ACTIVITY: Sucrose + ((2,6)-beta-D-fructosyl)(N) =

glucose + ((2,6)-beta-D-fructosyl)(N+1).

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- PTM: THE N-TERMINAL IS BLOCKED.

CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: L41732: AAB36606.1: -
 DR InterPro: IPR003469: Glyco_hydro_68.
 DR Pfam: PF02435: Glyco_hydro_68; 1.
 KW Transferrase; Glycosyltransferase; Signal.
 FT SIGNAL 1 30
 FT PROPEP 31 51
 FT CHAIN 52 584
 FT CHAIN 584 584 LEVANSUCRASE.
 SQ SEQUENCE 584 AA: 63525 MW: CBBFA139AD0B8CE CRC64;

Query Match 11.5%; Score 281.5; DB 1; Length 584;
 Best Local Similarity 26.4%; Pred. No. 5.8e-11;
 Matches 120; Conservative 59; Mismatches 184; Indels 91; Gaps 26;

5 DYKEEYGFANHTRAIMKI-----PGQNSP-QKRVQFNAS-AIKINDSAKDYK 53
 79 DPQSDP-FARWTRADALDIKAHSDATVAAGONSIPQALTPMIPADFEVIN----- 128
 54 SGNLIDLVMSWPQONADGTAAVHGHIYSALAGDPK-----NSDDTPLH-----LFYOKV 106
 129 -----PDVWVMDPTWTLIDKHADQFSTNGWEVIFCLTADNAGYGFPRDRIYHARIGFYRA 184
 107 GDTSTI-----DSKNAAGVEED--MDKFPVNDPLKQTOEMSGSATLTR--DGOVRLFY 157
 185 GIPASRRPVNGGWTPGCHLFPDGASAOVYACQTY--NQAEWSGSSRLQIHGNTVSFVY 242
 158 TDYSGNRPEDGGGAGN-----QLISTAOVNLSDPAATLKVDSVSDHKSVDGDDGTVYON 213
 243 TDVAFNRD--ANANNITRPAALITQTLGRIHADENHWFTEFAHTPLLO-PDGLVLQN 298
 214 IQQFIDECKWISGDNHTLDRPHYVEDKCH--KYLVEFANT-----GTTDGYOGDOSFNNK 266
 299 GAQ-----NEFFNRDPTEFTEDPKHPGVNYVVEGNTAGQGVANCTEADLGFR-- 347
 267 AYYGSDVFPQNEKN-KLLQSPKQOIASLANGALGIVELADDTYK--SVNKPPLVASTV 323
 348 -----PNDPNAETLQEVLDGCAVYOKANIGLA-IATDSTLSKKKPLSLISANCY 396
 324 ADEVERANIFKNNKMYLFTDSRSGKMTSDGINDKD-VYMLGPGDLSLNGPHNPIN-ETG 381
 397 NDQTERPQVYLLNGKYYVIFTISHRTFAA-GVDDGQGY--GFVDDGIRSDPQPNVYSG 453
 382 LVLANMLDPADLTHTYSHCGIPHEGNNVLTSTY 415
 454 LTMG-----NPTDLNTAAGTDPDPSPDQNRARQSY 484

RESULT 8

INVB_ZYMO STANDARD: PRT; 413 AA.

AC 060115; 060117; 060125;
 DT 15-DEC-1998 (Rel. 37; Created)
 DT 15-DEC-1998 (Rel. 37; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Extracellular sucrose (EC 3.2.1.26) (Beta-fructofuranosidase)
 DE (Invertase) (Saccharase) (Protein B46).
 GN SACC OR INVB OR SUCB3.
 OS Zymomonas mobilis.
 CC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
 CC Zymomonas.
 CC NCBI_TaxID=542;

QY 172 GNOIISTAOVNLSDPATLTKVDG---VSDH-----KSYEDGDDGTYYONIOFI 218
 Db 169 -----PGAAIAKVRGRIVTSDGVGLKDFQYKRLFE-ADGITYOTENO-- 211
 QY 219 DEGWISGDNHTLDPHYVEDKGHKYLVFEANTGTTDGYOGDOSFNKKAYGSDVFEQN 278
 Db 212 -NSSMNRDPSFIDP-----NDGKLYWFFEGNVA---GERGHT-----VGAELGPVP 257
 QY 279 EKNLLSPKQIATSLANGALGIVELADDTYK--SVMKPLVASTVADYERANIFKM 336
 Db 258 PGHEETGARGPQV-----GCGLA-VAKDLSGEEMELLPLVTVAVGNDOTERPHVFOQ 311
 QY 337 NKWYLFDSRGSKMT-SDGINDKD-VYMLGPGDLSNGPHNPINETGLVLMNLDPADLT 394
 Db 312 GKTYLFTISH-KFTYADGLTGPQGY--GFVEHLPFGPRPMNASGLVLGN--PPQOP 365
 QY 395 HTYSHCGIPHEGNNVLTSM 416
 Db 366 QTYSHCVMP-----NGLVTSFI 382

RESULT 10

SACB_PSESG STANDARD: PRT: 415 AA.
 ID SACB_PSESG
 AC 052408;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Levanucrase (EC 2.4.1.10) (beta-D-fructofuranosyl transferase)
 DE (Sucrose 6-fructosyl transferase).
 GN LSC.
 OS Pseudomonas syringae (pv. glycinea).
 OC Bacteria: Proteobacteria; gamma subdivision: Pseudomonadaceae;
 CC Pseudomonas
 OX NCBI_TaxID=318;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=PG4180;
 RA MEDLINE=98394981; PubMed=9726857;
 RT Hettwer U., Jaekel F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.:
 "Cloning, nucleotide sequence, and expression in *Escherichia coli* of
 levanucrase genes from the plant pathogens *Pseudomonas syringae* pv.
 RT glycinea and *P. syringae* pv. phaseolicola.";
 RL Appl. Environ. Microbiol. 64:3180-3187(1998).
 CC -1- CATALYTIC ACTIVITY: Sucrose + ((2,6)-beta-D-fructosyl)(N) =
 CC glucose + ((2,6)-beta-D-fructosyl)(N+1).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF037443; AAC36056.1;
 DR InterPro: IPR003469; Glyco_hydro_68.
 DR Pfam: PF02435; Glyco_hydro_68; 1.
 DR Transferase; Glycosyltransferase.
 KW
 SQ SEQUENCE 415 AA; 45844 MW; 839B686AC80610CF CRC64;

Query Match 11.0%; Score 269; DB 1; Length 415;
 Best Local Similarity 25.1%; Pred. No. 2.4e-10;
 Matches 111; Conservative 66; Mismatches 136; Indels 130; Gaps 24;

QY 16 TRADMLKIPGQNSPOFVKPQFNASAIKINDSAGYKSGNLDLDVWDSWFLONAGTA 75
 Db 12 SRADALKV--NENDPTTQPLVSPDPFVMSDF-----VFMDTFLFLDLDTGV 57
 QY 76 ANGHYIVVALAGDPKNSDDTPLHL-----FYQVGDTSID 112

Db 58 VSVNGMSVIVTLTAD-RHPDD-PQYVGANGRYDIKRDMEHRHGRARMCWYSRTG---K 111
 QY 113 SMKAGVFEEDMKFVENDPLKQTOBMSGASATLTND-GVRLFTYDYSGNPDGCTGA 171
 Db 112 DWIEGGRVMAE-----GVSPTRBMAGTPLVLLNDKDDIDLYXCVCV----- 152
 QY 172 GNOIISTAOVNLSDPATLTKVDG-----VSDHKSVEFDGDDGTYYONIOFI 218
 Db 153 -----PGAAIAKVRGRIVTSDGVGLKDFTEVKTLFE-ADGKYOTENO-- 195
 QY 219 DEGWISGDNHTLDPHYVEDKGHKYLVFEANTGTTDGYOGDOSFNKKAYGSDVFEQN 278
 Db 196 -NSTWNRDPSFIDP-----NDGKLYWFFEGNVA---GERGHT-----VGAELGPVP 241
 QY 279 EKNLLSPKQIATSLANGALGIV--ELADDTYKSVMKPLVASTVADYERANIFKM 335
 Db 242 PGHEETGARGPQV-----GCGLAVAKDLSGDEM-BILPLVAVGNDOTERPHVFOQ 294
 QY 336 NKWYLFDSRGSKMT-SDGINDKD-VYMLGPGDLSNGPHNPINETGLVLMNLDPADL 393
 Db 295 DGKTYLFTISH-KFTYADGVTGPQGY--GFVEHLPFGPRPMNASGLVLGN--PPAOP 348
 QY 394 THYSHCGIPHEGNNVLTSM 416
 Db 349 FQTYSHCVMP-----NGLVTSFI 366

RESULT 11

SACB_ERWAM STANDARD: PRT: 415 AA.
 ID SACB_ERWAM
 AC 046654;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Levanucrase (EC 2.4.1.10) (beta-D-fructofuranosyl transferase)
 DE (Sucrose 6-fructosyl transferase).
 GN LSC.
 OS Erwinia amylovora.
 OC Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae;
 CC Erwinia.
 OX NCBI_TaxID=552;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EA/74;
 RA Geier G., Gelder K.K.:
 RT "Characterization and influence on virulence of the levanucrase gene
 RT from the fireblight pathogen *Erwinia amylovora*.";
 RL Physiol. Mol. Plant Pathol. 42:387-404(1993).
 CC -1- CATALYTIC ACTIVITY: Sucrose + ((2,6)-beta-D-fructosyl)(N) =
 CC glucose + ((2,6)-beta-D-fructosyl)(N+1).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X75079; CAAS2972.1;
 DR InterPro: IPR003469; Glyco_hydro_68.
 DR Pfam: PF02435; Glyco_hydro_68; 1.
 DR Transferase; Glycosyltransferase.
 KW
 SQ SEQUENCE 415 AA; 46408 MW; 4FF564F6E0607FEB CRC64;

Query Match 10.4%; Score 255; DB 1; Length 415;
 Best Local Similarity 25.4%; Pred. No. 1.9e-09;
 Matches 115; Conservative 64; Mismatches 141; Indels 132; Gaps 29;

QY 9 DYGFANI--TRADMLKIPGQNSPOFVKPQFNASAIKINDSAGYKSGNLDLD----- 61

```

Db      3  YNYPPLTMTRADALKV- HEDDPPTTQO-----VIDIAPVMS 39
QY      62  ----VWDSMPLONAGCTAANYGHYHVSALADPKMSDTPH----- 100
Db      40  EEVFLTMDPLMEDEFGEIISVNGWCIIFLTPLAD-RNTDNPFOFEDENGNYDITRDWEDRRG 98
QY      101  ----LFYOKVADTSDSWKNRNGAFVEDBDRKVPNDPFLKYOTQOEWSSGATLEKD-GOYR 154
Db      99  RARICWYWSRTG-----KDMIFGGRWV-AEGVAPT-----TREMACTPILINDGDID 145
QY      155  LFYTYDSGNPEDEGTGA--GNQIISTAOVNLSDPADATLKVDGVSDBKSVFDDGSDGTYY 211
Db      146  LYTTCVT-----PGATIAKRGKIVTSDQSVSL-----EGFOQVTSLSF-SADGTYI 190
QY      212  QNIQOFIDEGKNIISGDNHTLRDPHYEEDK-GHKLYVEFANTGTTPGYGDGDSFNKKAY 269
Db      191  QTEEO--NAFW-----NFRDPSPIEDNDKLYLMEGNAV--GPRGSHEI-TQAEW 237
QY      270  GGSVDYFONEKKLLQSPKQJASLANGALGYELADDDTVK--SVMKPLVASNTVADEV 327
Db      238  GNVPPEGYEDVGAKKQQA-----GCVGLA-VAKDLSGSMQILLPLITAVGYNDOY 286
QY      328  ERANIFKNNKKWYLEFTDSRGSKMT--SDGINDXK-VYMLPGGDSLNGPHNPINETGLVY- 384
Db      287  ERPHFVFODGKKYLLFTTISH-KYTFADNLTGCPDGYV--GFVSDKLTGPYTPMNSSGLVIG 342
QY      385  MNMLNADPDLTHTYSHCGIHPHECNAVVLTSYM 416
Db      343  NPSSOP---FQYTSYHVP-----NGLVYSFI 366

```

Accession	Source	Size	Score	YMMO	Standard	Prnt	423 AA
ID	SACB_ZYMMO						
AC	Q60114; Q60116; Q06487;						
DT	15-DEC-1998 (Rel. 37, Created)						
DT	15-DEC-1998 (Rel. 37, Last sequence update)						
DT	16-OCT-2001 (Rel. 40, Last annotation update)						
DE	Levansucrase (EC 2.4.1.10) (beta-D-fructofuranosyl transferase)						
DE	(Sucrose 6-fructosyl transferase).						
OS	SACB OR LEVU OR SUCE2.						
OS	Zymomonas mobilis.						
OC	Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;						
OC	Zymomonas.						
OX	NCBI_TaxID=542;						
RN	(1)						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=ATCC 10988 / ZML;						
RX	MEDLINE=93305726; PubMed=8318541;						
RA	Song K.B., Joo H.K., Rhee S.K.;						
RT	"Nucleotide sequence of Levansucrase gene (levu) of Zymomonas mobilis						
RL	ZML (ATCC10988).";						
RN	Biochim. Biophys. Acta 1173:320-324(1993).						
RL	(2)						
RN	SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.						
RP	STRAIN=IFO 13756;						
RC	MEDLINE=95218269; PubMed=7766026;						
RX	Kyono K., Yanase H., Tomomura K., Kawasaki H., Sekai T.;						
RA	"Cloning and characterization of Zymomonas mobilis genes encoding						
RT	extracellular levansucrase and invertase.";						
RL	Biotechol. Biochem. 59:289-293(1995).						
RN	(3)						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=NRRL B806;						
RA	Gunasakratan P., Mukundan G., Kannan R., Velmurugan S.;						
RT	Alt-Adelkader N., Alvarez E., Baratti J.;						
RL	Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.						
RN	(4)						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=ATCC 31821 / ZM4 / CP4;						
RA	Ahn J.Y., Kang H.S.;						
RT	"Sequence analysis of 44B6 fosmid clone of Zymomonas mobilis ZM4.";						
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.						

```

CC      -I- CATABOLIC ACTIVITY: Sucrose + ((2,6)-beta-D-fucosyl1)(N) =
CC      glucose + ((2,6)-beta-D-fucosyl1)(N+1).
CC      -I- SUBCELLULAR LOCATION: Secreted.
CC      -I- PM: DOES NOT SEEM TO BE N-TERMINALLY PROCESSED.
CC      -I- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: AF081588; AAA27695.1; -.
DR      EMBL: D17524; BAA04475.1; -.
DR      EMBL: L33402; AAA27702.1; -.
DR      EMBL: AF313764; AAG29870.1; -.
DR      InterPro: IPR003469; Glyco_hydro_68.
DR      Pfam: PF02435; Glyco_hydro_68; 1.
DR      KW      Transferase; Glycosyltransferase.
FT      CONFLICT   39      39      V -> I (IN REF. 2).
FT      CONFLICT   200     203     NPED -> TPKI (IN REF. 1 AND 4).
FT      CONFLICT   217     217     E -> Q (IN REF. 2).
FT      CONFLICT   220     220     T -> A (IN REF. 2).
FT      CONFLICT   244     244     C -> Y (IN REF. 2).
FT      CONFLICT   379     379     I -> V (IN REF. 2).
SQ      SEQUENCE   423 AA: 46762 MW: 19A4691DA3BB9FE3 CRC64;
```

Query Match	9.9%: Score 243; DB 1; Length 423;
Best Local Similarity	24.4%: Pred. No. 1, 1e-08;
Matches	108; Conservative 60; Mismatches 135; Indels 140; Gaps 25;
OY	16 TRADMLKIPGQNSPQRFVQFNASAIKIDSAGKDGKGNLIDL-----VWDSW 66
DB	14 TRADAMK--HTDPTATP-----TIDYDPVMTDKYKVVWDTW 50
OY	67 PLQADGTAANYHGYHIVSALAGD-----PKNSDPTPLHLYQKGDTSIDSKNNAGRV 120
DB	51 PLRQINGVVSFOQWSYIFALVADRTKYGMHNRNDGARIGIFYSRGG-----SNWIFGHL 106
OY	121 FEDMDKFPVNDPYLYKOTQEWSSGATLTKDG--QVRLFTDYSGNPEDCGTGAGNOIIS 177
DB	107 LKD-----GAMP-----RSMEWSG--CTIMAPGTANSVEVFFFSVNDTP----- 143
OY	178 TAOVNLSPDAPATLKVQGVSPDKKSV-FDG-----GDCTVQNLQOQFIDECKWISGD 227
DB	144 -----SESVAOCKGKGYTADKRSWMDGPKVKTDLQADGLYAD--YAENNFW---- 190
OY	228 NHTLRDPHYV-EDKGHKYLVFEANTGTTDYGQDSQFNKKAYYGGSDVFQNEKNKLQ 285
DB	191 --DPRDPHFVFNPDGQKTYALFEENV-----AMRGTVAAVGEIEIGPV- 231
OY	286 SPKROI--ASLANGALGIVE-LADDYTVASVYMKPLVASMTVDEVERANIFKNNKAYL 341
DB	232 PPKETEPDGAATYCAAIGIAQALAEATTEMKLRLPPLTATGAVNDQOTRPHVYVQNGLTYL 291
OY	342 FTDSRSGKMTSDGINDKDYVNLGFGG-----DSLNGPHNPINETGVLV-NMNLDPADL 393
DB	292 FTISHSHSYV-ADGLS-----GPDGYGVFSENGICGPEYPLGSLGLVGNPSSQP--- 340
OY	394 THHTYSHGCIHPREGNNAVLTSLYM 416
DB	341 YQAYSHYV-----TNGLVTSFI 358

RESULT 13

ID	BIGA_SALTY	STANDARD:	PRT:	1953 AA.
AC	P25927;	P25928;	O9XCO3.	
DT	01-MAY-1992	(Rel. 22, Created)		
DT	15-JUN-2002	(Rel. 41, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		

FT	CONFLICT	1698	1698	D -> N (IN REF. 1)
FT	CONFLICT	1795	1798	SA -> T (IN REF. 1).
FT	CONFLICT	1836	1837	O -> T (IN REF. 1).
SEQ	SEQUENCE	1953	AA; 200150 MW; 611B3F1C9549D91AE CRC64;	
Query Match				
Best Local Similarity 5.7%; Score 139; DB 1; Length 1953;				
Matches 109; Conservative 59; Mismatches 199; Indels 142; Gaps 23;				
QY	54	SGNLIDLDVW-----DSWPLQNDAGTAANYHGXYHSALADGPKNSDDPFLHLHLYOKVG	107	
Db	278	NGSVLEGAAGVWNYSHQDNQWOLTTADGKTLANTGMDVDYDANAAVIEGTEGNLLYKKYDSRG	337	
QY	108	DTSIDSKMKNAGVEEDMKREVPNDPYLYKQTOEMSGS-----TLTKDQVRLFYTD	159	
Db	338	YLIIAD--DNTTVISGDDQAHNSRGMDSIGQDRTGYIISGDRYVNTLTGSSV-----TD	391	
QY	160	YSGNPEEGCGTCAGQIIISTAQV-----LSQPPAAT	190	
Db	392	GATGMVVISGDDTNTFTISGHSIVDANGALISGNGTTTNFAGDIASVGGGTAAIIGDNAT	451	
QY	191	LKYDVGSDHK-----SVFDGCGDTGYVQNIQOFIDE--GKVISGDNHTLRDPH-----	235	
Db	452	IKNIGTSDISAGSTGYIDGNMARNVNDGIMTITTDGSGHITGDNVYIDNAGSTYVG	511	
QY	236	-----YVEDKGKRYLF-EANTGTTCGYOGDOSFNKKAY-----YGSDFEFQ	277	
Db	512	ADATAYLIE--GDNALVINEGNIITSGAVGTRIDGDHAHTTNGDIADVAGSAVAIIN	569	
QY	278	NEKKKLLQSPKKQJASLANCALGIVELADDTYKSVKPLVASNTVADEV-----ERANI	332	
Db	570	GDNGSLTQAGDLV--TDGAMGIITTCGTGMEARNTGATVRD--ADSVGFVVAAGEKNT	623	
QY	333	FKANNKMYLFTDRCG-----SKMTSDG-----INDKDYVMLGPGGSLNGPNPI	377	
Db	624	FK--NKGDIDVSLNGTGAALVSGDMSQVTLDDGINVYSQDSEGYFSSATGVSVSGDSNAV	681	
QY	378	NETGLVLMNLDPADLHTYVSHCIP-----HPEGNNVLTYSM-----TNRGFY	422	
Db	682	DITG---NVNIS-ADYGDDDLAAGAPPLTGVVVGNGNTVTLNGALNIDDDLSATGQY	737	
QY	423	PEHSHLRDLKGVNIKSGDT-----SGCEN	447	
Db	738	-----LDVVGLSVTGDDNDVEIDGGIN	759	
RESULT 14				
GTF1_STRDO STANDARD; PRT: 1597 AA.				
ID	GTF1_STRDO	STANDARD;	PRT: 1597	AA.
AC	P11001;			
DT	01-JUL-1989 (Rel. 11, Created)			
DT	01-JUL-1989 (Rel. 11, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)			
DE	(Sucrose 6-glucosyltransferase).			
GN	GTFI.			
OS	Streptococcus downei (Streptococcus sobrinus).			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1317;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MEP28;			
RX	MEDLINE=87308014; Pubmed=3040686;			
RA	Ferretti J.J., Gilpin M.L., Russell R.R.B.;			
RT	"Nucleotide sequence of a glucosyltransferase gene from Streptococcus			
RT	sobrinus MFE28."			
RL	J. Bacteriol. 169:4271-4278(1987).			
CC	-I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS THAT ARE THOUGHT			
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE			
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE			
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.			
CC	-I- CATALYTIC ACTIVITY: Sucrose + [(1,6)-alpha-D-glucosyl](N) = D-			

```
CC fructose + ((1,6)-alpha-D-glucosyl)(N+1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M17391; AAC63063.1;
CC InterPro: IPR002479; CW_binding.
CC InterPro: IPR003318; GH_70.
CC Pfam; PF01473; CW_binding_1; 19.
CC Pfam; PF03324; Glyco_hydro_70; 1.
CC Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
CC SIGNAL 38
CC CHAIN 39 1597
CC DOMAIN 39 1050 CATALYTIC (APPROXIMATE).
CC DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).
CC DOMAIN 1099 1597
CC REPEAT 1099 1132 1.25 A, 2 B, AND 5 AC REPEATS.
CC REPEAT 1163 1213 A REPEAT.
CC REPEAT 1227 1277 AC REPEAT.
CC REPEAT 1292 1342 AC REPEAT.
CC REPEAT 1352 1399 B REPEAT.
CC REPEAT 1406 1455 AC REPEAT.
CC REPEAT 1465 1512 B REPEAT.
CC REPEAT 1519 1568 AC REPEAT.
CC REPEAT 1582 1597 A REPEAT (INCOMPLETE).
CC SEQUENCE 1597 AA; 177080 MW; B9E86A20868798E CRC64;
-----
Query Match 5.5%; Score 135; DB 1; Length 1597;
Best Local Similarity 22.9%; Pred. No. 0.53;
Matches 80; Conservative 30; Mismatches 111; Indels 128; Gaps 19;
-----
QY 36 QPNASAIKNIDSAQYDKSGNMLIDLVDSWPLONADG-----TAANYGHYHISA-- 86
DB 1209 QHNQNAVNTNTVA---DKTGH-----WYLGKDSQVAVTGAQTYGKQHLFYEPANQQ 1256
QY 87 LAGDPKNSDPTPLHLFYOKVGDTSIDSKN-----AGRVFEDMKREVPNDP----- 132
DB 1257 VKGDFVFAKDGKLYFYDVDSG---DMWTNFTFIEDKAGNMF-----YLGKGAAVTGAQT 1307
QY 133 -----YLYKQTOEMSGSATITLKDGQVRLFTYDYSNGNEDDGTGAGNIIITAVYNLSQP 186
DB 1308 IKGKGLFKFANGQOQKGVIVKADCKIRYV-----DAQT--GEQVEN----- 1347
QY 187 DAATLKYDVGSDHKSVPEDGSGTYVONI-----OOFFID-----EGKMTSGDNHTLR 232
DB 1348 ---KVSVNG---KTYFGSGDTAQTQANPKGQTFKSSQVLRFRYNLEGOVSSGWTET 1401
QY 233 DPH-----YVEDKGH--KYLVEFANTGTT---DGYQGDQSF-- 263
DB 1402 AEHEMVVYVSKGVLTAQTIGNQRYEFKDNQHOVKGGLVGTGNDGKLYRYDANSQDQAFNK 1461
QY 264 -----NNKAYVYGGDVFPQNEKNKLLQSPKQIADSLANALGIVELADY 308
DB 1462 SVTVNGKTYFYFGSDGTAQTQAN-----PKGQTFKDSGVLRFYNLEGOY 1505
-----
RESULT 15
GTFB_STRMU STANDARD: PRT: 1476 AA.
AC P08987; O69381; O69384; O69387; O69390; O69396;
```

```
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN GTFB.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN-GS-5;
CC MEDLINE=87308013; PubMed=3040685;
CC Shiroza T., Ueda S., Kuramitsu H.K.;
CC "Sequence analysis of the gtfB gene from Streptococcus mutans.";
CC J. Bacteriol. 169:4263-4270(1987).
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN-MT4239, MT4245, MT4467, and MT8148;
CC MEDLINE=98231643; PubMed=9570124;
CC Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
CC Kimura S., Hamada S.;
CC "Molecular analyses of glucosyltransferase genes among strains of
CC Streptococcus mutans.";
CC FEMS Microbiol. Lett. 161:331-336(1998).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) = D-
CC fructose + ((1,6)-alpha-D-glucosyl)(N+1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M17361; AAA88588.1;
CC EMBL; D88651; BAA26101.1;
CC EMBL; D88654; BAA26105.1;
CC EMBL; D88657; BAA26109.1;
CC EMBL; D88660; BAA26113.1;
CC EMBL; D89977; BAA26119.1;
CC PIR; B33135; B33135.
CC InterPro: IPR002479; CW_binding.
CC InterPro: IPR003318; GH_70.
CC Pfam; PF01473; CW_binding_1; 13.
CC Pfam; PF03324; Glyco_hydro_70; 1.
CC Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
CC SIGNAL 34
CC CHAIN 35 1476
CC DOMAIN 35 1051
CC DOMAIN 1097 1476
CC REPEAT 1097 1130
CC DOMAIN 1161 1470
CC REPEAT 1161 1210
CC REPEAT 1225 1275
CC REPEAT 1290 1340
CC REPEAT 1355 1405
CC REPEAT 1420 1470
CC VARIANT 62
-----
S -> T (IN STRAIN MT4239).
```

```
FT VARIANT 65 65 I -> T (IN STRAINS MT4239, MT4245,
FT VARIANT 68 68 MT4251, MT4467 AND MT8148).
FT VARIANT 78 78 A -> V (IN STRAIN MT4239).
FT VARIANT 86 86 O -> P (IN STRAIN MT4251).
FT VARIANT 89 89 S -> I (IN STRAIN MT4239).
FT VARIANT 168 168 S -> F (IN STRAIN MT4251).
FT VARIANT 168 168 K -> N (IN STRAIN MT4251).
FT VARIANT 276 276 D -> S (IN STRAINS MT4239, MT4245 AND
MT4251).
FT VARIANT 399 399 N -> R (IN STRAIN MT4239).
FT VARIANT 474 474 I -> T (IN STRAIN MT4239).
FT VARIANT 512 512 K -> R (IN STRAIN MT8148).
FT VARIANT 519 519 F -> Y (IN STRAIN MT8148).
FT VARIANT 701 701 T -> I (IN STRAIN MT8148).
FT VARIANT 708 708 A -> V (IN STRAIN MT8148).
FT VARIANT 938 938 F -> L (IN STRAIN MT8148).
FT VARIANT 952 952 YGTPVA -> FGKPE (IN STRAINS MT4245,
MT4251 AND MT8148).
FT VARIANT 963 964 MT -> SV (IN STRAINS MT4245, MT4251 AND
MT8148).
FT VARIANT 968 970 VDG -> ADS (IN STRAINS MT4245, MT4251 AND
MT8148).
FT VARIANT 1086 1086 A -> T (IN STRAIN MT4239).
FT VARIANT 1158 1158 S -> N (IN STRAIN MT4239).
FT VARIANT 1163 1163 H -> Y (IN STRAIN MT4251).
FT VARIANT 1168 1168 E -> K (IN STRAIN MT8148).
FT VARIANT 1182 1182 Y -> C (IN STRAIN MT8148).
FT VARIANT 1234 1234 A -> P (IN STRAIN MT4239).
FT VARIANT 1263 1263 H -> R (IN STRAIN MT8148).
FT VARIANT 1263 1263 H -> R (IN STRAINS MT4239, MT4245 AND
MT4251).
FT VARIANT 1264 1264 H -> Y (IN STRAINS MT4245 AND MT4251).
FT VARIANT 1272 1272 G -> S (IN STRAINS MT4245 AND MT4251).
FT VARIANT 1329 1329 Y -> H (IN STRAINS MT4239, MT4245, MT4251
AND MT8148).
FT VARIANT 1394 1394 H -> Y (IN STRAINS MT4245 AND MT4251).
FT VARIANT 1402 1402 G -> S (IN STRAINS MT4245 AND MT4251).
FT VARIANT 1459 1459 Y -> H (IN STRAIN MT4467).
FT CONFLICT 570 570 R -> A (IN REF. 1).
FT CONFLICT 800 817 ADOOVRVAASPSTGK -> LIKMFALRLARHQQA
(IN REF. 1).
FT CONFLICT 1310 1310 H -> L (IN REF. 1).
SQ SEQUENCE 1476 AA: 165685 MW: 3479B62B07694D98 CRC64;
```

Query Match 5.5%; Score 134; DB 1; Length 1476;
Best Local Similarity 21.3%; Pred. No. 0.56; Indels 135; Gaps 21;

Matches 80; Conservative 47; Mismatches 113; Indels 135; Gaps 21;

```
QY 68 LQNDGTANYHG-----YHVSALAGDPKNSDDPTPLHLFYOKVGPFSIDSWKNA 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1132 LKNEDGTYA--YGNDRREYNGYQFMGVRHFNCE-----MSVGLTVID----- 1177
QY 118 GRV--FEDMDKRVNDPLYKQTOEMSGSATLLTKDGOVRLFTYDSGN----- 163
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1178 GQVOYFDEM-----YQAK--GKFTVTADGKIR-YFDKQSGNMVRRRIENEEG 1223
QY 164 -----PEDGGTAGNOIISTAOVNLSDPAATLKVGDSDHK--SVFDGGDGTVYONI 215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1224 KMLYLGEDGAALVYGSOTIN-GQHLFRANGVQVKGFEVTDHGRISYYDGSNGDOIRN-- 1280
QY 216 QFT--DEGKMTSGDNH-----TLRDPHY-----VEDKHKYLVFEANTGTTDGYOG 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1281 RFVFNAGOWEYFDNNGYAVTGARTINGQHLFRANGVQVKGFEVTDYGRISYYDGSNG 1340
QY 260 DQ-----SFNNKAY-----YGSDFEFOKKNKLLOSPKCIASLANGA 298
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1341 DOJRNRFVRNAOGWEYFDNNGYAVTGARTINGQHLFR-----ANG- 1382
QY 299 LGIYELADDTYKSVKPLVASNTVADEVERANIFKMNKNWLYLFTDSRGSKMTSDGINDK 358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1383 ---VQVKGFEVTDHGRISYYDGSNGDOIRNRFVRNAOGWEYF-DNNGYAVT----- 1431
QY 359 DVTWLGPGGDSLNGPH 374
```

Db 1432 -----GARTINGQH 1440

Search completed: June 16, 2003, 12:07:27
Job time : 30.9037 secs


```

Db 93 DWMSPLOMADGTAAHYHYSALAGDPKNSDTPHLFTYOKVGDTSIDSMNAGRV 152
Qy 121 FEDMKFVNDPYLKYTOEMSGSATLTRKDGVRLEFTYDYSGNPDGAGNOIISTAQ 180
Db 133 FEDMKFVNDPYLKYTOEMSGSATLTRKDGVRLEFTYDYSGNPDGAGNOIISTAQ 212
Qy 181 VMLSOPDAATLKVDGVSDBKSVFDGDDGTYYONIOQFIDEKMWISGDNHTLRDPHYEDK 240
Db 213 VMLSOPDAATLKVDGVSDBKSVFDGDDGTYYONIOQFIDEKMWISGDNHTLRDPHYEDK 272
Qy 241 GHKYLVEFANTGTTCGYOGDGSFNNKAYGSDVFEONKKNLLOSPKKQIASLANGALG 300
Db 273 GHKYLVEFANTGTTCGYOGDGSFNNKAYGSDVFEONKKNLLOSPKKQIASLANGALG 332
Qy 301 IVELADDTYKSVKPLVASTNTVADEVERANIFKNNKNWYLFETDSRGSKMTSDGINDKV 360
Db 333 IVELADDTYKSVKPLVASTNTVADEVERANIFKNNKNWYLFETDSRGSKMTSDGINDKV 392
Qy 361 YMLGFGDSLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHEGNNVLTSTYMTNRG 420
Db 393 YMLGFGDSLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHEGNNVLTSTYMTNRG 452
Qy 421 FYPEHSHLRDKLGVNKGSDTSGGENSSGOGQF 455
Db 453 FYPEHSHLRDKLGVNKGSDTSGGENSSGOGQF 487

```

RESULT 2

```

ID 097181 PRELIMINARY: PRT: 489 AA.
AC 097181.
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE levensucrase.
GN CAC1772.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng O.,
RA Gibson R., Lee H.M., Dubois J., Olu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucetite-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007686; AAK7937.1;
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
KM Complete proteome.
SQ SEQUENCE 489 AA; 54553 MW; 9002CB364FID3CD8 CRC64;

```

Query Match 71.4%; Score 1752; DB 16; Length 489;
 Best Local Similarity 71.1%; Pred. No. 2,2e-102;
 Matches 323; Conservative 51; Mismatches 78; Indels 2; Gaps 2;

```

Qy 2 NSGDYKEDYGFATIRADMLKIPGOONSPOKRVPOFNASAIKNIDSAGYKSGNLIIDL 61
Db 30 NDMNKETTYGVSHITRYNMSKIPMEONDLKFKVPEFNASTLKNIASAGYKSGNLIIDL 89
Qy 62 WMDSPFLONADGTAAHYHYSALAGDPKNSDTPHLFTYOKVGDTSIDSMNAGRV 121
Db 90 WMDSPFLONADGTAAHYHYSALAGDPKNSDTPHLFTYOKVGDTSIDSMNAGRV 149
Qy 122 EDMKFEVNDPYLKYTOEMSGSATLTRKDGVRLEFTYDYSGNPDGAGNOIISTAQ 181
Db 150 KDSKRYVANDPYLKYTOEMSGSATLTRKDGVRLEFTYDYSGNPDGAGNOIISTAQ 209

```

```

Qy 182 NLSOPDAATLKVDGVSDBKSVFDGDDGTYYONIOQFIDEKMWISGDNHTLRDPHYEDK 241
Db 210 NLSOPDSMTINIDSDHKSVFDGDDGTYYONIOQFIDEKMWISGDNHTLRDPHYEDK 269
Qy 242 HKLYLVEFANTGTTCGYOGDGSFNNKAYGSDVFEONKKNLLOSPKKQIASLANGALG 300
Db 270 HKLYLVEFANTGTTCGYOGDGSFNNKAYGSDVFEONKKNLLOSPKKQIASLANGALG 329
Qy 301 IVELADDTYKSVKPLVASTNTVADEVERANIFKNNKNWYLFETDSRGSKMTSDGINDKV 360
Db 330 IVELADDTYKSVKPLVASTNTVADEVERANIFKNNKNWYLFETDSRGSKMTSDGINDKV 389
Qy 361 YMLGFGDSLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHEGNNVLTSTYMTNRG 420
Db 390 YMLGFGDSLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHEGNNVLTSTYMTNRG 449
Qy 421 FYPEHSHLRDKLGVNKGSDTSGGENSSGOGQ 453
Db 450 MYSDBHSSFPDFFLNLINGKTSTVSNSTLQOQ 483

```

RESULT 3

```

ID 0925E5 PRELIMINARY: PRT: 499 AA.
AC 0925E5.
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE levensucrase.
GN SACH.
OS Paenibacillus polymyxa (Bacillus polymyxa).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=1406;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CF43;
RX MEDLINE=21042000; PubMed=11200435;
RA Bezotte S., Aymerich S., Chamber R., Czarne S., Berge O., Heulin T.;
RT "Isolation of the Paenibacillus polymyxa levensucrase gene impairs
RT its ability to aggregate soil in the wheat rhizosphere.";
RL Environ. Microbiol. 2:333-342(2000).
DR EMBL: AJ133737; CAB39327.1;
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
SQ SEQUENCE 499 AA; 55317 MW; E646CD986292336C CRC64;

```

Query Match 58.8%; Score 1443; DB 2; Length 499;
 Best Local Similarity 61.2%; Pred. No. 6.3e-83;
 Matches 278; Conservative 54; Mismatches 114; Indels 8; Gaps 4;

```

Qy 2 NSGDYKEDYGFATIRADMLKIPGOONSPOKRVPOFNASAIKNIDSAGYKSGNLIIDL 61
Db 33 DASDLKTNAPFQITIRNMLNFKQGEKEKEVSEFSDSTIKNIPASGRDSSGKLIFD 92
Qy 62 WMDSPFLONADGTAAHYHYSALAGDPKNSDTPHLFTYOKVGDTSIDSMNAGRV 121
Db 93 WMDSPFLONADGTAAHYHYSALAGDPKNSDTPHLFTYOKVGDTSIDSMNAGRV 152
Qy 122 EDMKFEVNDPYLKYTOEMSGSATLTRKDGVRLEFTYDYSGNPDGAGNOIISTAQ 180
Db 153 KDNOKLNLANDPILKNOSEWMSGSATLTRKDGVRLEFTYDYSGNPDGAGNOIISTAQ 210
Qy 181 VMLSOPDAATLKVDGVSDBKSVFDGDDGTYYONIOQFIDEKMWISGDNHTLRDPHYEDK 240
Db 211 INVSPDDTKLIDVDEBLKSTYDGDGKTYONVOOSG---VDMNHTLRDPHYEDK 266
Qy 241 GHKYLVEFANTGTTCGYOGDGSFNNKAYGSDVFEONKKNLLOSPKKQIASLANGALG 300
Db 267 GHKYLVEFANTGTTCGYOGDGSFNNKAYGSDVFEONKKNLLOSPKKQIASLANGALG 326
Qy 301 IVELADDTYKSVKPLVASTNTVADEVERANIFKNNKNWYLFETDSRGSKMTSDGINDKV 360

```


Db 327 IVELNDYTLKNVMPPLIASLVYDEITERANVFKMGLMYLFTSNGSKYTVDAIGDDI 386
OY 361 YMLPGGDSLNGPNHPINETGGLVLMNLDPADLTHYSHGCIPEEGNNVLTSTMTNRG 420
Db 387 YMLGYVSTSLTGPKPLNGTGLVLIHODLRDVTWTAHFAIPGKNNVVSSTMTNRG 446
OY 421 FYPEHSHLRDLKLVNTKGSPTSGENS-SGOG 453
Db 447 LEPDKSTFAPSFLNKGSKTSYVKNGLIEGQ 480

RESULT 4

O9AEX9 PRELIMINARY: PRT: 415 AA.
AC O9AEX9: 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)
DE Levanusucrase.
OS Rahmella aquatilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Rahmella.
OX NCBI_TaxID=34038;
RN [1]
RP SEQUENCE FROM N.A.
RA Kim H., Cha J.;
RT "Cloning and characterization of levanusucrase";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY027657; AAK14794.1;
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1.
SQ SEQUENCE 415 AA; 45958 MW; F09F5A1BF8690C5E CRC64;

Query Match 12.2%; Score 300; DB 2; Length 415;
Best Local Similarity 26.7%; Pred. No. 3.9e-11;
Matches 116; Conservative 60; Mismatches 150; Indels 108; Gaps 22;

OY 16 TRADMLKIPGQNSPQKVPQFNASAIKINIDSAKGYKSGNLIIDVDSWPLQADGTA 75
Db 12 TRDALAV--NENPTTQPIVSADFPYMSD-----EYFIMDAFPLNSLDGTV 57
OY 76 ANYGCHIVSALADGPKNSDDTPLH-----LEYQVGDTSIDSW 114
Db 58 VSVGWSVITFLTAQRNNNNSEYLDADGNYDTSDMNNRHRGAKICWYSRTG---KDW 113
OY 115 KNAARVEDMDKFPNDPYLKQTOEMSGS-ATLTKGQVPLFTYDSGNPDGCTGAGN 173
Db 114 IFGRVMAE-----GVSPTTREVAGTPIILLNEDEVDLYTTCVT-----PGA 155
OY 174 QIISTAQNLSQPDATLKVDSVSDHKSVEFDGDTVQNTQOFIDEKWTISGDNHLLRD 233
Db 156 TIAVRGKVLTSSEGVTLA--GENEVKSLF-SADGVYIQTESQ---NTYNNFRDPSPIID 209
OY 234 PHYEDKGKHYLEANTGTTDGYOGQDSFNKKAYYGGSDVFQNEKNKLLQSPKKQIAS 293
Db 210 PH---DGKLYMVEGVNA-----GER-----GSHVIGKQEMGTL--PRDHRDVG 248
OY 294 LANGALIVELA-----DDYTVKSVMKPLVASNTVADEVERANIEFMNKKVILFTDSR 346
Db 249 NARQACIGMAVAKDLSDGEM---EILPLVTAVGVNDQTEREPEVFODKKYLLFTLISH 305
OY 347 GSKMT-SDGINDKD-VYMLPGGDSLNGPNHPINETGGLVLMNLDPADLTHYSHGCI 404
Db 306 --KFTYAEGLTGPQGV--GFLSDRLTGTPSPMNGSLVLCN--PPSQPQIYSHCWMP- 358
OY 405 PEGNNVLTSTMTN 418
Db 359 ----NGLVTSEFIDN 368

RESULT 5
O97179 PRELIMINARY: PRT: 428 AA.

AC O97179;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Levanusucrase.
GN CAC1774.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hiltl J., Wolf Y.I.,
RA Tatusov R.L., Sabatne F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007666; AAK79739.1;
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1.
KM Complete proteome.
SQ SEQUENCE 428 AA; 48395 MW; 9670B154B178E23E CRC64;

Query Match 12.0%; Score 295; DB 16; Length 428;
Best Local Similarity 27.7%; Pred. No. 8.4e-11;
Matches 112; Conservative 62; Mismatches 147; Indels 84; Gaps 21;

OY 59 DLDVDSWPLQADGTANANYGHIVSAL-----AGDKNSDDPLHLYQKVGDTSDS 113
Db 66 NLWVDTPLVKKKDSGLAVNMGKYIFALTSRNVGMNKRHDVAGISYFCSTDG-----EN 121
OY 114 WKNARVEEDMDKFPNDPYLKQTOEMSGSATLTKGQVPLFTYDSGNPDGCTGAGN 173
Db 122 WYKGLAVNVEDAL-----GSRQWAGSALIDENGMVQFFYT-----ATGRKG 163
OY 174 QIISTAQNLSQPDATLKVDSVSDHKSVEFDGDTVQNTQOFIDEKWTISGDNHLLRD 233
Db 164 EAVRFEQRLVTKFSINVDKGVHITNCSKHQVILE--PDGVYIQTMOQ--AKGPII--- 217
OY 228 NHTLBDHYVEDKHK--YLVEANTGTTDGYOGQDSFNKKAYYGGSDVFQNEKNKLLQ 285
Db 218 -YSFDPYFFEDPKTKKYLIFEGNKGGKIEKMPENIGDLF-----RKNHIAIP 266
OY 286 SPKKQIASLANGALCI-VELADYTVKSVMKPLVASNTVADEVERANIEFMNKKVILFTD 344
Db 267 RGVENF---NGNVGIAVAOKKDLRFKLLPLLEAVGVNQDLERQIVMKKKYLLFTI 322
OY 345 SRGSMT-SDGINDKQVYMLPGGDSLNGPNHPINETGGLV--NMNLDPADLTHYSHGCI 402
Db 323 SH--KFTYAEGLTGPQGV--GFLSDRLTGTPSPMNGSLVLCN--PPSQPQIYSHCWMP- 358
OY 403 PHEGNNVLTSTMTNRGTFPEHH--SHLR-----DKLVNIGK 439
Db 377 ---SGHDV-----SFINEYHFNQGLRIGGTAPTLQISLKG 410

RESULT 6
O9EVD6 PRELIMINARY: PRT: 630 AA.
AC O9EVD6: 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Fructosyltransferase.
GN FTF.
OS Actinomyces naeslundii.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Actinomycetaceae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1655;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 12104.
 RA Bergeron L.J., Moron-Bermudez E., Burne R.A.;
 RT "Characterization of the fructosyltransferase gene of *Actinomyces*
 naesiundii";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF228582; AAC09737.1; -
 DR InterPro; IPR003469; Glyco_hydro_68.
 DR Pfam; PF02435; Glyco_hydro_68; 1.
 KW Transferase.
 SQ SEQUENCE 630 AA; 68256 MW; CCA04184DD6D7D2 CRC64;

Query Match 11.9%; Score 291; DB 2; Length 630;
 Best Local Similarity 25.3%; Pred. No. 2.6e-10;
 Matches 133; Conservative 72; Mismatches 208; Indels 112; Gaps 29;

QY 7 KEDYGFHITRADMIKT-----PQONSPOFKVPOFNASAIKNIDSAGYDKSGNLIDL 60
 DB 120 EENY-TAKMTRADAMQIOIRVFQPOGAVRAPRLPEQLMPEISN-----GEPATSE--DV 171
 QY 61 DWDSNPLOMDGTANTHGHYISALAGDPK---NSDDTPLH---LFYQKV----- 107
 DB 172 WWDWTWLTDEAHQISYNGWEIAFSLVADRHAGYTFDDRHARLGFYRKAGTQSSA 231
 QY 108 ---DTSIDSMKNAGVFEDMDKFNPNDEYLKYQTOEMSGSATLTKDGOVRLFTYDYGNP 164
 DB 232 DGAGSSNGWYIGGHVFPDAS--ARSSRTYHAQTEMGSGARLMGKNKIRNFYTSYST 289
 QY 165 EDGCTG-AGNOIISTAQVNLSP-DAAATLKYDGVSDHKSVPDGGDGYVONI---OQFI 218
 DB 290 APPRGTTHQVHAT-----SKPYDPRIVQSEG---RIYATNGVLTGFEPTQHOLLV 340
 QY 219 DEGKWIS-----GDNHTLRDPHYVEDKGH-----KYLEFANTGTTDYGQDGSFNKA 267
 DB 341 PDGKTYOTREONPGVN--FRDPFTFRDONNPSDLETWVEFGNS---AFVREQIYDAA 394
 QY 268 YYGSGDVEFFON-----EKNKLQSPKKQIASLANGALGIVELAD-----DYTV 310
 DB 395 AKAGQNTLATCTEEDLEGYEGK---DPKAEVNAVNGRGYQLANGLARAKNKAMTE 450
 QY 311 KSVMPPLVASTVADDEVBRANIEFKNNKMYLFTSRGSKMTSDGINDKD--VYMLGPGGDS 369
 DB 451 WEYLPRLSLGNCVNDQTERPQIYHODGKYLLFTISH--REYADGLQGEVGY--GFLVGDG 507
 QY 370 LNPENPINE-TGLVL-----NMNLDPAD-----LFTYSHGCIHPPEGNNV--V 411
 DB 508 LRSQKPLNOMTGIALGNPILNPNPKPKYSPDENQSPYTTQSTSHYMP---GGLVESF 564
 QY 412 LSTYTNRGFYPEHHSHLRDLKLVNIKSGDTSGE--NSSGQGF 454
 DB 565 IDSIGNKDGNPVRGSLSPVTKLINSIGDTSVDRYCTGNLGLGF 609

RESULT 7

Q93TMI PRELIMINARY; PRT; 431 AA.
 AC Q93TMI;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Levanusucrase LscC (EC 2.4.1.10).
 GN LSCC.
 OS *Pseudomonas syringae* (pv. *glycinae*).
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas.
 OX NCBI_TaxID=318;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PG4180;
 RX MEDLINE=98394981; Pubmed=9726857;
 RA Heltner U., Jaechel F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.;
 RT "Cloning, nucleotide sequence, and expression in *Escherichia coli* of

RT Levanusucrase genes from the plant pathogens *Pseudomonas syringae* pv.
 RT *glycinae* and *P. syringae* pv. *phaseolicola*.
 RL Appl. Environ. Microbiol. 64:3180-3187(1998).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-PG4180;
 RX MEDLINE=21242726; Pubmed=11344135;
 RA Li H., Ullrich M.S.;
 RT "Characterization and Mutational Analysis of Three Allelic *lsc* Genes
 RT Encoding Levanusucrase in *Pseudomonas syringae*.
 RL J. Bacteriol. 183:3282-3292(2001).
 DR EMBL; AF346402; AAK49952.1; -
 DR InterPro; IPR003469; Glyco_hydro_68.
 DR Pfam; PF02435; Glyco_hydro_68; 1.
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 431 AA; 47619 MW; 0C71D0F133071AA5 CRC64;

Query Match 11.5%; Score 283; DB 2; Length 431;
 Best Local Similarity 26.2%; Pred. No. 4.8e-10;
 Matches 116; Conservative 63; Mismatches 135; Indels 128; Gaps 25;

QY 16 TRADMKTIGQONSPOFKVPOFNASAIKNIDSAGYDKSGNLIDLDWDSNPLOMDGTA 75
 DB 28 SRADALKV--NENDPTTQPLVSADFPVNSDT-----VFIMDTMPLRLDGTIV 73
 QY 76 ANYGHYISALAGDPKNSDDTPLHL-----FYQVGDTSID 112
 DB 74 VSNAGSVILTLFAD-RHEDD-PQYLDANGRIDKRDWEDRGRARMCYWSRTG-----K 127
 QY 113 SWKNAGVFEDMDKFNPNDEYLKYQTOEMSGSATLTKD-GQVRLFTYDYGNEPDGTGA 171
 DB 128 DWIFGCAVMAE-----GVSPTREWAGPILINDGDIDLVTCT----- 168
 QY 172 GNOIISTAQVNLSPDAAATLKYDG---VSDH-----KSVPDGGDGYVYONIQFI 218
 DB 169 -----PAAAVAKVGRIVISDQVELKDTQVAKLE-ADGTYQTEAQ-- 211
 QY 219 DEGKWISGDNHTLRDPHYVEDKGHKLYFEAVNTGTTDYGQDGSFNKAYGSDVFFON 278
 DB 212 -NSSMNPDPSPPTIDP-----NDGKLYMVEGVNA---GERSHT-----VGAALGVP 257
 QY 279 EKKKLQSPKKQIASLANGALGIVELADDTYVK--SVKPLVASTVADDEVBRANIEFKN 336
 DB 258 PGHEIDVGARFGY-----GCIGILA-VAKDLSEEMELLPLVTAGVNDQTERPHYVQD 311
 QY 337 NKMYLFTDSRGSKMT--SDGINDKD--VYMLGPGGDSLNGPHNINFTGLVNMNLDPADLT 394
 DB 312 GKTYLFTISH--KFTYADGITGPDGY--GVGEHLFGFYRPMNASGLVLR--PPEQPF 365
 QY 395 HTYSHGCIHPPEGNNVLTYSY 416
 DB 366 QFYSHCVMP-----NGLVTSFI 382

RESULT 8

Q93FUG PRELIMINARY; PRT; 424 AA.
 AC Q93FUG;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Levanusucrase.
 GN LSCA.
 OS *Pseudomonas aurantiflora*.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas.
 OX NCBI_TaxID=86192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S-4380;
 RA Koh I., Jang E.K., Kim J.S., Song K.S., Kim C.H., Rhee S.K.;
 RT "Screening, molecular cloning and characterization of a novel
 levanusucrase, and its expression in *Escherichia coli*.";

Query Match	11.3%	Score 278.5;	DB 2;	Length 578;
Best Local Similarity	23.7%;	Pred. No. 1.4e-09;		
Matches 125;	Conservative 67;	Mismatches 193;	Indels 143;	Gaps 25

[illegible]

DB 60 ITGETVKNEDWYVMALVANRADTGDTEGWHNRNFAVIGFYYSRNG---ID-WTFGGR 115
 QY 120 WEEDMDKFPNDPRLKYQTOEWSGATL---TKDQVRLFYEDVSGNPEEDGTGAGNOII 176
 DB 116 LIQKADKLRPD-----EWSGLVIRAGTKN-TIDPEFTSVN----- 150
 QY 177 STAQVNLQOPDAATLKVDGVSDDKSVFDG-----GDGTVYONIQ--QFIDEGKMTS 225
 DB 151 --TDINOSVPSGCTKIFS-NDEKVMFEGFSKTVEMFMSADGVNANAEEDGYFD----- 201
 QY 226 GDNHRLRPHYED--KGKHYLFEPANGTGTGYGDSFNNK-----AYY 269
 DB 202 ----FRDPHPLNADPKIKYCLFE--GNMPGVGKFTLHDAEIGAVPPGYTPAGAOY 253
 QY 270 GGSDFVQNEKNKLQSPKKQIASLANGALGVELADYTVK-----SYMKPLVASN 321
 DB 254 GAA-----AIGIARLADGAYEKGDSSRHHWTLLPALVAL 288
 QY 322 TVADEVERANIFKMNKKYLFETDSRSGSKTSDGINDKDVY-MLGPGGDSLNGPHNPINET 380
 DB 289 GVNQDTERPHYVFKNNTYIFTISHSTYTGDSGTGPDGYGVFSENG--IFGYPBPLNAS 346
 QY 381 GLVLMNNDPADLTH-TYSHGCIHPREG 407
 DB 347 GLVLDG--NPSAPYETYSH--FYDDPG 369

RESULT 11

Q932F7 PRELIMINARY: PRT: 1141 AA.

AC 0932F7
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding
 DE protein.
 GN SDRF OR SAV0563.
 OS Staphylococcus aureus (strain Mu50./ ATCC 700699).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Staphylococcus.
 OX NCBI_TaxId=158878;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21311952; PubMed-11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iken J.-O., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Iabuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 aureus";
 RT Lancet 357:1225-1240(2001).
 DR EMBL, AP003355; BAB56725.1.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR TIGRFAMs: TIGR01167; LPTG_anchor; 1.
 DR TIGRFAMs: TIGR01168; YSIRK_signal; 1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Complete proteome.
 SO SEQUENCE 1141 AA; 124038 MW; E679F7C2991846D9 CRC64;

Query Match 5.7%; Score 141; DB 16; Length 1141;

Best Local Similarity 21.7%; Pred. No. 1.6; Mismatches 198; Indels 170; Gaps 26;

Matches 119; Conservative 61; Mismatches 198; Indels 170; Gaps 26;
 QY 3 SGDKEDYGFAMHTRADM-----KIPQONSPOKFPQFN-----ASAINKIDSAR 49
 DB 464 AGSQVDYGNIKLNGSTIIDONTEIKYKVNPNQQLPQSNIDYDFSGYEDVTSGDFMK 523
 QY 50 GYDSGNLIDDVMD-----SWPLQVNDGTA-----NTHGCHI 83

DB 524 SF--SNNVATLDFGINSAYIIKVVSKYPTISDGLDIAQGTSMRTTKGYGYNAGYSN 581
 QY 84 VSALAGDKNSDDT--PLHLEFYQKGDPTSIDSKNAGVFPDMKFPVNDPYLKYQTOEW 141
 DB 582 FIVYSNDGGGGGTAKPEPKLY-KIGD-----YWEDEVDR---DVOGTDSKEK 626
 QY 142 SGS---ATLT--KDQVRLFYTDYSGNPEEDGTGAGNOIISTAQVNLQOPDA-ATLKVDGV 196
 DB 627 PMANLVLTLPYDGTGKTSVRDANANGHEFGGLKDS---EYTVKFEPTAGYLPKVNKT 682
 QY 197 SDKSVFEDGGGTGYQNIQ-----FIDEGKMTSGNHLRDPHYEDKGHKLYVEA 249
 DB 683 TDGEKDSNGSSITVYKINCKDMSLDTGFKYKPKYMLGD-----YWEEDTNKDGIDDA 734
 QY 250 NTGTTDGYOGQGSFNKKAAYGSDVFPQNEKNKLQSPKKQIASLANGALGVELAD--- 306
 DB 735 NEP-----GIKDV-----KYLKDSIGKAYIGTTTDSASKYKFTDLDN 772
 QY 307 -DYTVK-----SYMKPLVASNTVAD-----VERANIEKNN-----K 338
 DB 773 GNYTVFEFTPAGYPTVKNNTAEDKDSNGLTGTVIKDADNMTLDSGFYKTPKYSLDGY 832
 QY 339 WYLFDSGSKMTSGINDKDVYMLGPGGDSLNGPHNPINETGLVLMNLD----- 389.
 DB 833 WYDSNKKDKQDSTEGIDVAVTLLNEKEVIG--TTKTDNGKRYFPDNLDSGKRYKIFE 890
 QY 390 -PADLFTTYSHC--GIPPEGNNVLT-----SYMTNGRFYPEHSHLRDLKGVNIKSD 441
 DB 891 KPAGLTQTVTNTEDKRDADGGEVDVTTTDDHDFLDNGYFEE-----D 934
 QY 442 TSGGENSS 449
 DB 935 TSDSDSDS 942

RESULT 12

Q9F3Z5 PRELIMINARY: PRT: 1449 AA.

AC Q9F3Z5
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Adhesion penetration protein.
 GN App.
 OS Neisseria meningitidis.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxId=487;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-23515;
 RA Ala-Aldeen D.A.;
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-23515;
 RA Abdel Hadi H., Woolridge K.G., Robinson K., Ala Aldeen D.A.;
 RT "Identification and Characterisation of App: an Immunogenic
 RT Auto-transporter Protein of Neisseria meningitidis";
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AJ296276; CAC14670.1.
 DR MEROPS: S06.006;
 DR InterPro: IPR000710; IGA_S6.
 DR InterPro: IPR004899; Pertact_sup.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR Pfam: PF03212; Pertactin; 1.
 DR Pfam: PF03212; Pertactin; 1.
 DR PRINTS: PR00921; IGASERPTASE.
 DR SMART: SM00020; Tryp_Spc; 1.
 DR PROSITE: PS00135; TRYP_SIN_SER; UNKNOWN_1.
 KW Hydrolase; Serine protease.
 SO SEQUENCE 1449 AA; 159071 MW; C07850BD4B7C25D CRC64;

Query Match 5.7%; Score 141; DB 2; Length 1449;

Best Local Similarity 22.4%; Pred. No. 2.2;
Matches 110; Conservative 68; Mismatches 193; Indels 120; Gaps 27;
OY 3 SGD-YKEDGFAHITRADMKIPGQNSPOFKVPOFNASKIKNDKSGDKSG----- 55
DB 239 SGDRHNDYCPMPPLAGA-----AGDSGSPFIYDKTNKKMLNGVLQGTGYPSGREGNF 293
OY 56 NLIDLVDWDSW-----PLQNA-D-GTAANYGHYIVSAL--ACDPKNSD 95
DB 294 QLIKRD-----WFYDDIRGOTHTVFEPRSGHFSFTSNNGTGTVETNEKVSNPRLKY 349
OY 96 DTPHLHFLYQKVGDTSIDSKNAGVFEEDMKFVP-----NDPYLKYOTQEMSGSATLT 149
DB 350 QT-VRLFDESINETDKPEPVYAAGV-----NOYRPLNNGENLSEFDYNGKRLIISNNINQ 404
OY 150 DGOVRLFTYDYSGNPEDEGTGAGNQIISTAOVNLSDPAATLKVDYSDHKSVEFGDGT 209
DB 405 GAGGLYFEGDFTVSPENNETWOG-----AGVHISEDSTVWKVNGVANDR-LSKIGKGT 457
OY 210 VYQNIQPFIDEKMWISGDNHTLRDPHYVEDKHKYLFEE-----ANTGTTDGYOGDQSFN- 264
DB 458 LHVQAKG-ENOGSISVGDGTIYLD-QQADDKGKQAFSEIGLVSGRGTVQ-LMNDQENP 514
OY 265 NKAYYG-----GSDVFFQNEKKNKLLQSPKKQIASLANGALGIVELADYTVKSVK 315
DB 515 DKLYFGRGRGLDNGHLSLFSH-----RIQNTDEGAMIV-----NHNAATTST 557
OY 316 PLVANSNTVADVERANIKFMN-----NKWYLFTDSGSKMTSDGIN-----DKDYVM 362
DB 558 VTIIGNESITQPSGKNIRLNSKEIAYNGMF---GEQDTTKTKGRRLVYQPAEDRTL 614
OY 363 LGPGDGLNGPHNPINETGLVLMNMLDPADLTHTYSHC-----GIPHEGNNVLT 414
DB 615 LLSGGTNLNG---NITQTNKLFESGRPT--PHAYNHLGSGWKMKEGIPGE---IYWDN 666
OY 415 YMTNRGFLPEH 425
DB 667 DWINRTFKAEH 677
RESULT 13
O9JMB4 PRELIMINARY; PRT; 1449 AA.
AC O9JMB4;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE Iga-specific serine endopeptidase (EC 3.4.21.72).
GN IGA2 OR NNA0457.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCB1_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
RA Jagers K., Leather S., Mould S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds K., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491";
RL Nature 404:502-506(2000).
EMBL, AL162753; CAB83754.1;
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; pertact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; pertactin; 1.
DR PRINTS: PRO0921; ICASERPTASE.
DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.

KW Hydrolase; Complete proteome.
SQ SEQUENCE 1449 AA; 159070 MW; B858DB54BE50146 CRC64;
Query Match 5.7%; Score 141; DB 16; Length 1449;
Best Local Similarity 22.4%; Pred. No. 2.2;
Matches 110; Conservative 68; Mismatches 193; Indels 120; Gaps 27;
OY 3 SGD-YKEDGFAHITRADMKIPGQNSPOFKVPOFNASKIKNDKSGDKSG----- 55
DB 239 SGDRHNDYCPMPPLAGA-----AGDSGSPFIYDKTNKKMLNGVLQGTGYPSGREGNF 293
OY 56 NLIDLVDWDSW-----PLQNA-D-GTAANYGHYIVSAL--ACDPKNSD 95
DB 294 QLIKRD-----WFYDDIRGOTHTVFEPRSGHFSFTSNNGTGTVETNEKVSNPRLKY 349
OY 96 DTPHLHFLYQKVGDTSIDSKNAGVFEEDMKFVP-----NDPYLKYOTQEMSGSATLT 149
DB 350 QT-VRLFDESINETDKPEPVYAAGV-----NOYRPLNNGENLSEFDYNGKRLIISNNINQ 404
OY 150 DGOVRLFTYDYSGNPEDEGTGAGNQIISTAOVNLSDPAATLKVDYSDHKSVEFGDGT 209
DB 405 GAGGLYFEGDFTVSPENNETWOG-----AGVHISEDSTVWKVNGVANDR-LSKIGKGT 457
OY 210 VYQNIQPFIDEKMWISGDNHTLRDPHYVEDKHKYLFEE-----ANTGTTDGYOGDQSFN- 264
DB 458 LHVQAKG-ENOGSISVGDGTIYLD-QQADDKGKQAFSEIGLVSGRGTVQ-LMNDQENP 514
OY 265 NKAYYG-----GSDVFFQNEKKNKLLQSPKKQIASLANGALGIVELADYTVKSVK 315
DB 515 DKLYFGRGRGLDNGHLSLFSH-----RIQNTDEGAMIV-----NHNAATTST 557
OY 316 PLVANSNTVADVERANIKFMN-----NKWYLFTDSGSKMTSDGIN-----DKDYVM 362
DB 558 VTIIGNESITQPSGKNIRLNSKEIAYNGMF---GEQDTTKTKGRRLVYQPAEDRTL 614
OY 363 LGPGDGLNGPHNPINETGLVLMNMLDPADLTHTYSHC-----GIPHEGNNVLT 414
DB 615 LLSGGTNLNG---NITQTNKLFESGRPT--PHAYNHLGSGWKMKEGIPGE---IYWDN 666
OY 415 YMTNRGFLPEH 425
DB 667 DWINRTFKAEH 677
RESULT 14
O99W46 PRELIMINARY; PRT; 1141 AA.
AC O99W46;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE Ser-Asp rich fibrinogen-binding, bone stalo-protein-binding
DE protein.
GN SDRE OR SA0521.
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OX Staphylococcus.
OX NCB1_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146;
RX Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Itoh T., Kamamoto M.,
RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Uti Y.,
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA Hiraoka H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA Ogasawara N., Hayashi H., Hiratsuka K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus";
RL Lancet 357:1225-1240(2001).
EMBL, AP003131; BABA1752.1;
DR InterPro: IPR001899; Gram_pos_anchor.

DR Pfam: PF00746; Gram_pos_anchor: 1.
DR TIGRfams: TIGR01167; LPTG_anchor: 1.
DR TIGRfams: TIGR01168; YSTRK_anchor: 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1141 AA; 124026 MW; 44541900B8C5A4F8 CRC64;

Query Match 5.7%; Score 139; DB 16; Length 1141;
Best Local Similarity 21.7%; Pred. No. 2.1;
Matches 119; Conservative 60; Mismatches 199; Indels 170; Gaps 26;

QY 3 SGDKEDYGFHITRADML-----KIPGOQNSPOKVPQFN-----ASAINKIDSAR 49
DB 464 AGSQVDYDGNIKLNGSTIIIDONTETIKYKYNPQQLPQSNRIYDSQYEDVTSQFDKK 523
QY 50 GYDKSGNLIDLDVWD-----SMPLOADGTAA-----NYHGYH 83
DB 524 SF--SNVATLDFGDINSAYIIKYSKPTPSDGLDIAOGTSMRTDKYGYNYAGYSN 581
QY 84 VSALAGDPKNSDDT--PLHLFYQKVGDTSIDSMKNAGRVFEDMDKFPVNDPYLKYOTQEW 141
DB 582 FIYTSNDFTGGGDTYKPEEKLY-KIGD-----YVWEDVDK---DGVQGTDSKEK 626
QY 142 SGS---ALLT-KDQVRLFTYDYSGNPEDEGTAGNQIISTAQVNLSPDA-ATLKVDGV 196
DB 627 PMANVLVLTLYPDGTTKSVRTDANGHYEFGGLKDG---EYTVKFEETPGYLPRTKYNGT 682
QY 197 SDHKSVFDGDDGYVYQNIQO-----FIDEGKWIISDNHTLRDPHYEDKGHKYLVEA 249
DB 683 TDGEDNSGSSITVYKINGKDMSLDTGYKEPKYNLGD-----YVWEDTINKGIDDA 734
QY 250 NTGTDTYQGDQSFNNKAYVGGSDVFFQNEKNKLLQSPKKOIASLANGALGIVELAD-- 306
DB 735 NEP-----GIKDY-----KVLKDSGTGKAVIGTTTIDASGKFKFTDLN 772
QY 307 -DYTVK---SYMKPLVASTVADN-----VEGANIFKMN-----K 338
DB 773 GNYTVEFTPGAGYTPVYKNTTAEDKDSNGLTGTTGVIKADNMNTLDSGFKTPKYSLGDV 832
QY 339 WYLFDSRGSKMTSDGINDKDYMLPGGDSLNGPHNPINETGLVNNMLD----- 389
DB 833 WYDSKDKDQDSTENGKIDVYVTLNKEGEVIG--TTTDENGKRFNDLSGKRYLVE 890
QY 390 -PADLFTHTYSHC--GIRPEGNVVL-----SYMTNRGYPYEHSHLRDLKGVNIKSD 441
DB 891 KPAGLTQTYVTNTTEDDKADGGEVDVTTIDHDFTLNNGYFEE-----D 934
QY 442 TSGGENSS 449
DB 935 TSDSDSDS 942

RESULT 15

ID 086489 PRELIMINARY; PRT: 1166 AA.
AC 086489;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Sdr E protein.
GN SDR E.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEWMAN;
RX MEDLINE=99098700; PubMed=9884231;
RA Josefsson E., McCrea K., Ni Eidhin D., O'Connell D., Cox J., Hook M.,
RA Foster T.J.;
RT "Three new members of the serine-aspartate repeat protein multigene
family of Staphylococcus aureus.";

RL Microbiology 144:3387-3395(1998).
DR EMBL: AJ005647; CAA06652.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor: 1.
DR TIGRfams: TIGR01167; LPTG_anchor: 1.
DR TIGRfams: TIGR01168; YSTRK_anchor: 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 1166 AA; 126548 MW; 750A7B0135287DA4 CRC64;

Query Match 5.6%; Score 138; DB 2; Length 1166;
Best Local Similarity 22.2%; Pred. No. 2.5;
Matches 122; Conservative 53; Mismatches 201; Indels 174; Gaps 28;

QY 3 SGDKEDYGFHITRADMLKIPGOQNSPO-----KVPQNSAIAINKIDSA 48
DB 469 AGSQVDYDGNIKLNGSTIIIDONTETIKYKYNPQQLPQSNRIYDSQYEDVTSQFDKK 527
QY 49 GYDKSGNLIDLDVWD-----SMPLOADGTAA-----NYHGYH 82
DB 528 KSF--SNVATLDFGDINSAYIIKYSKPTPSDGLDIAOGTSMRTDKYGYNYAGYS 585
QY 83 IVSALAGDPKNSDDT--PLHLFYQKVGDTSIDSMKNAGRVFEDMDKFPVNDPYLKYOTQEW 140
DB 586 FIYTSNDFTGGGDTYKPEEKLY-KIGD-----YVWEDVDK---DGVQGTDSKEK 630
QY 141 WSGS---ALLT-KDQVRLFTYDYSGNPEDEGTAGNQIISTAQVNLSPDA-ATLKVDG 195
DB 631 KPMANVLVLTLYPDGTTKSVRTDANGHYEFGGLKDG---EYTVKFEETPGYLPRTKYN 686
QY 196 VSDHKSVFDGDDGYVYQNIQO-----FIDEGKWIISDNHTLRDPHYEDKGHKYLVE 248
DB 687 TDGEDNSGSSITVYKINGKDMSLDTGYKEPKYNLGD-----YVWEDTINKGID 738
QY 249 ANTGTDTYQGDQSFNNKAYVGGSDVFFQNEKNKLLQSPKKOIASLANGALGIVELAD-- 306
DB 739 ANEP-----GIKDY-----KVLKDSGTGKAVIGTTTIDASGKFKFTDLN 776
QY 307 -DYTVK---SYMKPLVASTVADN-----VEGANIFKMN-----K 338
DB 777 GNYTVEFTPGAGYTPVYKNTTAEDKDSNGLTGTTGVIKADNMNTLDSGFKTPKYSLG 835
QY 339 --WYLFDSRGSKMTSDGINDKDYMLPGGDSLNGPHNPINETGLVNNMLD----- 389
DB 836 YWYDSKDKDQDSTENGKIDVYVTLNKEGEVIG--TTTDENGKRFNDLSGKRYLVE 893
QY 390 -PADLFTHTYSHC--GIRPEGNVVL-----SYMTNRGYPYEHSHLRDLKGVNIK 439
DB 894 FEKPAGLTQTYVTNTTEDDKADGGEVDVTTIDHDFTLNNGYFEE----- 938
QY 440 SDTSGGENSS 449
DB 939 -DTSDSDSDS 947

Search completed: June 16, 2003, 12:09:40
Job time : 117.682 secs